GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 23:04:17 ; Search time 4272.77 Seconds
(without alignments)
17041.707 Million cell updates/sec

Title: US-10-027-000-3
Perfect score: 2502
Sequence: 1 atggctgatattgatgttga.....attggtggtctggcgtgtaa 2502

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	427	17.1	4193	8	KFBGLUC	X05918 Kluyveromyc
2	406.2	16.2	38269	1	AF486581	AF486581 Unculture
3	393.6	15.7	39305	8	SPU33009	U33009 Schizosacch
4	393.6	15.7	41799	8	SPBC1683	AL355920 Schizosac
5	393.6	15.7	85837	8	SPU33010	U33010 Schizosacch
6	392.2	15.7	2793	8	AF329731	AF329731 Volvariel
7	350.4	14.0	3578	1	AF005277	AF005277 Cellulomo
8	343.4	13.7	3710	1	ATUCBG1	M59852 A.tumefacie
9	302.2	12.1	2636	8	ABI293760	AJ293760 Agaricus
c 10	270	10.8	110000	1	REU80928_2	Continuation (3 of
11	205	8.2	2601	6	AX489445	AX489445 Sequence
12	197.4	7.9	37445	1	SC8A6	AL031013 Streptomy
13	184.4	7.4	3241	6	AR173232	AR173232 Sequence
14	184.4	7.4	17665	1	AF521878	AF521878 Streptomy
c 15	178.2	7.1	2760	1	SEY14327	Y14327 Saccharopol
16	175.2	7.0	12441	1	AF079762	AF079762 Streptomy
17	161	6.4	2256	6	E11369	E11369 DNA encodin
18	157.2	6.3	2640	1	TNBGLB	Z77856 T.neapolita
19	156.8	6.3	3179	1	CEGDGB	D14068 C.gilvus ge
c 20	152.2	6.1	11763	1	AE005774	AE005774 Caulobact
21	149.4	6.0	14878	1	AE005849	AE005849 Caulobact
c 22	143.6	5.7	10615	1	AE012240	AE012240 Xanthomon
c 23	138.6	5.5	15052	1	AF055579	AF055579 Streptomy
c 24	138.6	5.5	50937	6	AR159871	AR159871 Sequence
c 25	137.4	5.5	21243	1	AE001690	AE001690 Thermotog
26	136.2	5.4	3265	1	RAU92808	U92808 Ruminococcu
27	135.8	5.4	2166	6	AR205082	AR205082 Sequence
28	132.6	5.3	3158	1	RALBGLT	X15415 Ruminococcu
c 29	129.2	5.2	2075	1	ARGERMA	M11276 Arthrobacte
30	127.6	5.1	2620	1	CSBGLZ	Z94045 C.stercorar
31	127.6	5.1	4537	1	AF090429	AF090429 Azospiril
32	126.4	5.1	2346	1	PAL276021	AJ276021 Prevotell
c 33	126	5.0	12140	1	AE011777	AE011777 Xanthomon
c 34	123.2	4.9	110000	2	LMFLCHR36_08	Continuation (9 of
35	118.2	4.7	3542	1	AB071166	AB071166 Gluconace
36	117.2	4.7	3955	1	MIBBGLA	L06134 Microbispor
c 37	114.4	4.6	11557	1	AE004598	AE004598 Pseudomon
38	111.6	4.5	15852	1	AE006929	AE006929 Mycobacte
39	111.6	4.5	36300	1	MTCI28	Z97050 Mycobacteri
40	109	4.4	13332	1	AE005788	AE005788 Caulobact
41	106.2	4.2	2370	1	PRU35425	U35425 Prevotella
42	104.4	4.2	2894	1	CTBGLB	X15644 Clostridium
c 43	104.4	4.2	12882	1	AE013715	AE013715 Yersinia
44	104.4	4.2	150754	9	AC023491	AC023491 Homo sapi
45	104.4	4.2	258050	1	AJ414153	AJ414153 Yersinia

ALIGNMENTS

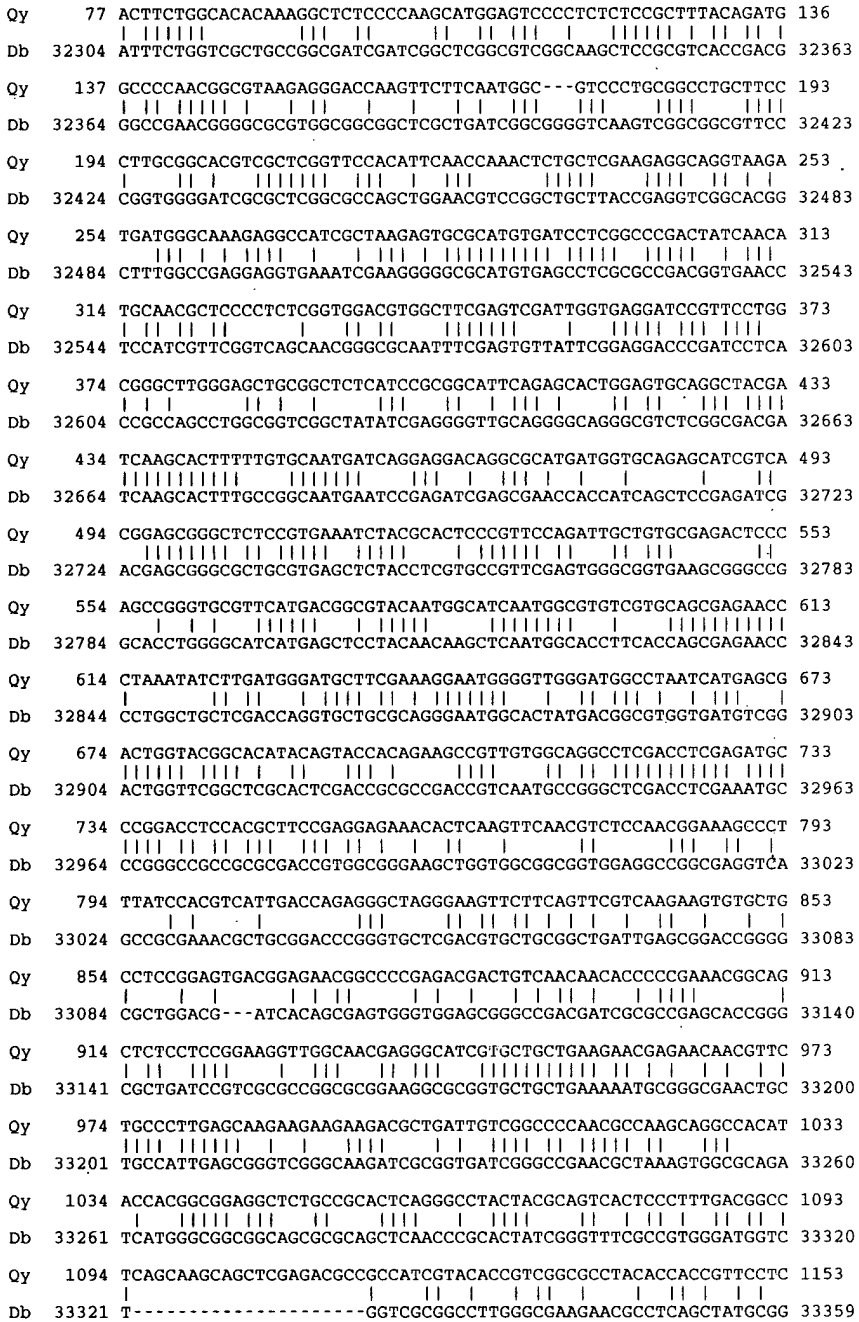
RESULT 1
KFBGLUC
LOCUS KFBGLUC 4193 bp DNA linear PLN 12-SEP-1993
DEFINITION Kluyveromyces fragilis beta-glucosidase gene.
ACCESSION X05918
VERSION X05918.1 GI:2804
KEYWORDS beta-glucosidase.
SOURCE Kluyveromyces marxianus.
ORGANISM Kluyveromyces marxianus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE 1 (bases 1 to 4193)
AUTHORS Raynal,A., Gerbaud,C., Francingues,M.C. and Guerineau,M.
TITLE Sequence and transcription of the beta-glucosidase gene of
Kluyveromyces fragilis cloned in Saccharomyces cerevisiae

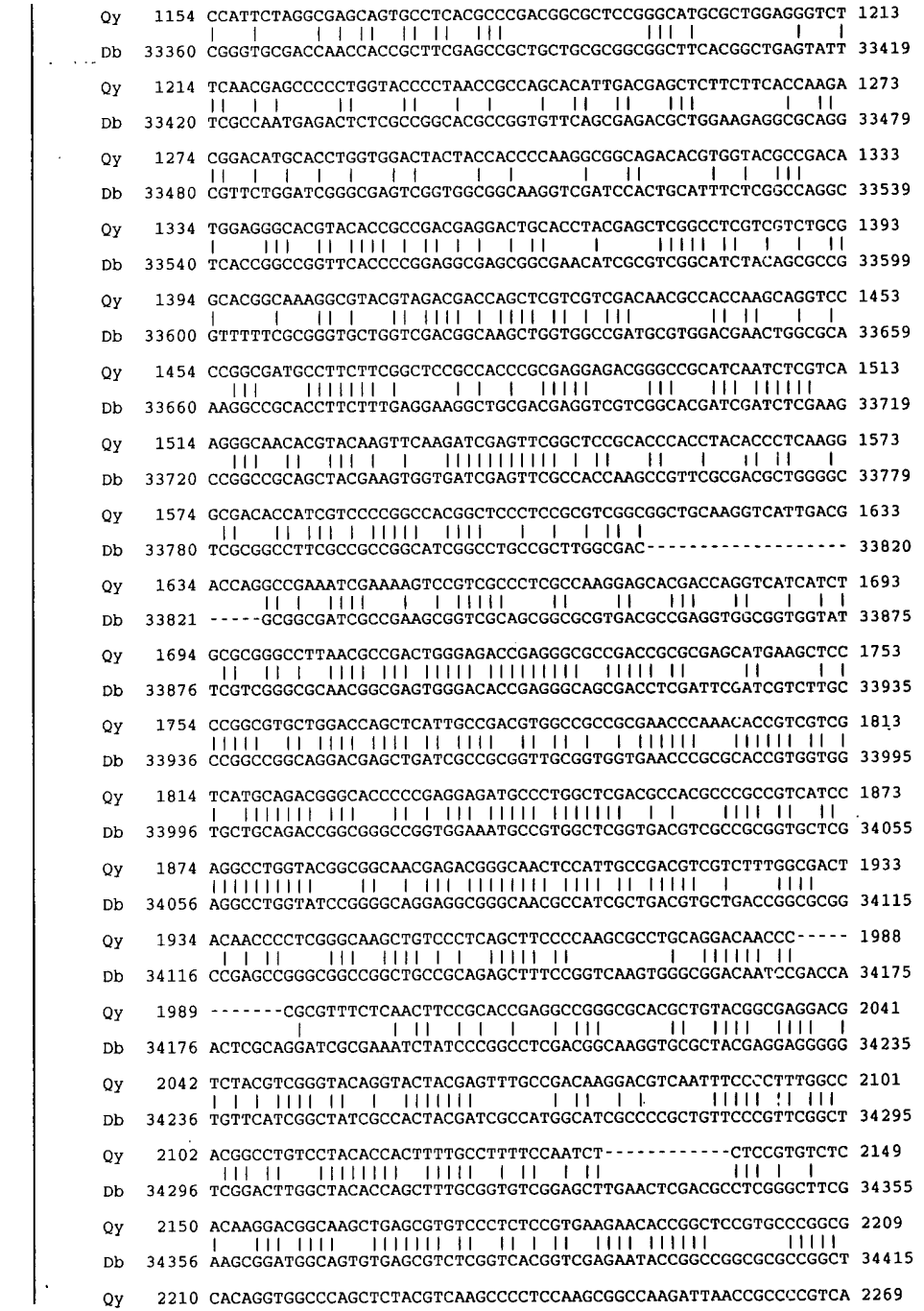
```
  JOURNAL   Curr. Genet. 12 (3), 175-184 (1987)
  MEDLINE   88210533
   PUBMED   2835179
FEATURES             Location/Qualifiers
     source          1. .4193
                     /organism="Kluyveromyces marxianus"
                     /strain="ATCC 12424"
                     /db_xref="taxon:4911"
                     /clone_lib="PHCG3"
     repeat_region   268. .276
                     /note="direct repeat 1"
     repeat_region   307. .315
                     /note="direct repeat 1"
     promoter        316. .320
                     /note="TATA-Box like sequence"
     promoter        342. .355
                     /note="TATA-Box like sequence"
     repeat_region   400. .408
                     /note="direct repeat 2"
     promoter        406. .412
                     /note="TATA-Box like sequence"
     repeat_region   416. .424
                     /note="direct repeat 2"
     misc_feature    459
                     /note="pot. transcription start site"
     misc_feature    482. .483
                     /note="pot. alternate transcription start site"
     CDS             612. .3149
                     /note="beta-glucosidase (AA 1 - 845)"
                     /codon_start=1
                     /protein_id="CAA29353.1"
                     /db_xref="GI:2805"
                     /db_xref="SWISS-PROT:P07337"
                     /translation="MSKFDVEQLLSELNQDEKISLLSAVDFWHTKKIERLGIPAVRVS
                     DGPNGIRGTKFFDGVPSGCFPNGTGLASTFDRDLLETAGKLMAKESIAKNAAVILGPT
                     TNMQRGPLGGRGFESFSEDPYLAGMATSSVVKGMQGEGIAATVKHFVCNDLEDQRFSS
                     NSIVSERALREIYLEPFRLAVKHANPVCIMTAYNKVNGDHCSQSKKLLIDILRDEWKW
                     DGMLMSDWFGTYTTAAAIKNGLDIEFPGPTRWRTRALVSHSLNSREQITTEDVDDRVR
                     QVLKMIKFVVDNLEKTGIVENGPESTSNNTKETSDLLREIAADSIVLLKNKNNYLTSK
                     ERRQYHVIGPNAKAKTSSGGGSASMNSYYVVSPYEGIVNKLGKEVDYTVGAYSHKSIG
                     GLAESSLIDAAKPADAENAGLIAKFYSNPVEERSEDEEPFHVTKVNRSNVHLFDFKHE
                     KVDPKNPYFFVTLTGQYVPQEDGDYIFSLQVYGSGLFYLNDELIIDQKHNQERGSFCF
                     GAGTKERTKKLTLKKGQVYNVRVEYGSGPTSGLVGEFGAGGFQAGVIKAIDDDEEIRN
                     AAELAAKHDKAVLIIGLNGEWETEGYDRENMDLPKRTNELVRAVLKANPNTVIVNQSG
                     TPVEFPWLEEANALVQAWYGGNELGNAIADVLYGDVVPNGKLSLSWPFKLQDNPAFLN
                     FKTEFGRVVYGEDIFVGYRYYEKLQRKVAFPFGYGLSYTTFELDISDFKVTDDKIDIS
                     VDVKNTGDKFAGSEVVQVYFSALNSKVSRPVKELKGFEKVHLEPGEKKTVNIELELKD
                     AISYFNEELGKWHVEAGEYLVSVGTSSDDILSVKEFKVEKDLYWKGL"
     repeat_unit     3314. .3322
                     /note="inverted repeat A"
     misc_feature    3442. .3445
                     /note="yeast termination consensus sequence"
     misc_feature    3472. .3474
                     /note="yeast termination consensus sequence"
     misc_feature    3498. .3501
                     /note="yeast termination consensus sequence"
     misc_feature    3541. .3543
                     /note="yeast termination consensus sequence"
     repeat_unit     3642. .3652
                     /note="inverted repeat B"
     repeat_unit     3655. .3666
                     /note="inverted repeat B'"
     repeat_unit     3714. .3722
                     /note="inverted repeat A'"
BASE COUNT     1338 a    675 c    941 g   1239 t
ORIGIN

  Query Match             17.1%;  Score 427;  DB 8;  Length 4193;
  Best Local Similarity   50.7%;  Pred. No. 1.5e-42;
  Matches 1253;  Conservative    0;  Mismatches 1155;  Indels   63;  Gaps    7;

Qy      1 ATGGCTGATATTGATGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGAT 60
          ||| || |  ||||||||||     | ||||   |  | | || ||  || ||| |

Db    612 ATGTCTAAATTTGATGTTGAACAGTTATTGAGTGAATTGAACCAGGATGAAAAGATTTCC 671

Qy     61 CTGCTGGCTGGTATCGACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCT 120
          | ||  |||   | || |||||||| || |||    |       |   ||| | ||  |

Db    672 TTACTTTCTGCAGTTGATTTCTGGCATACTAAGAAGATTGAACGGTTGGGAATTCCAGCG 731

Qy    121 CTCCGCTTTACAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCTTCAATGGCGTCCCT 180
          |  |  || | ||||| || || ||  | |||||||| ||||||||  |||| || |||

Db    732 GTGAGGGTTTCTGATGGTCCAAATGGTATTAGAGGGACGAAGTTCTTTGATGGGGTTCCT 791

Qy    181 GCGGCCTGCTTCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAA 240
          | |  || ||||||   || ||   | | |  || || ||  | |    |||||| ||

Db    792 TCAGGATGTTTCCCTAATGGTACCGGGTTGGCATCTACTTTTGATCGCGATCTGCTTGAG 851

Qy    241 GAGGCAGGTAAGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGC 300
            |||||||| ||||||  || ||  | || || |||| |||   ||||||  | ||

Db    852 ACAGCAGGTAAGTTGATGGCTAAGGAATCGATTGCGAAGAATGCTGCTGTGATTTTGGGT 911

Qy    301 CCGACTATCAACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAG 360
          || || |  ||||||||||||    ||| | ||||| ||||| || || ||  |   |||

Db    912 CCAACCACAAACATGCAACGTGGTCCTTTGGGTGGTCGTGGTTTTGAATCATTCTCTGAA 971

Qy    361 GATCCGTTCCTGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGA 420
          |||||  |  || || ||  |||   || |  || |  |    || || ||  ||   ||

Db    972 GATCCATATCTTGCTGGTATGGCTACTTCTTCTGTTGTTAAAGGTATGCAAGGCGAAGGT 1031

Qy    421 GTGCAGGCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATGATGGTG 480
          |    |||||  | ||||| ||| | || || ||  ||| |||  ||  |      |

Db   1032 ATTGCTGCTACCGTTAAGCATTTTGTTTGTAACGACTTGGAAGACCAACGTTTTTCTTCG 1091

Qy    481 CAGAGCATCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATTGCT 540
          |   || ||  | ||  |||||||  | ||||| |||       || |||    | ||

Db   1092 AACTCAATTGTTTCTGAAAGGGCTCTGAGAGAAATTTACTTGGAGCCCTTCAGATTGGCA 1151

Qy    541 GTGCGAGACTCCCAGCCGGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCGTGTCG 600
          ||   | |  || | || | |    | ||||| || || ||     |||||||||

Db   1152 GTTAAACATGCCAATCCTGTTTGTATAATGACTGCTTATAACAAGGTCAATGGCGATCAT 1211

Qy    601 TGCAGCGAGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGC 660
          |||  | |   |   ||      |    |  ||  |  || | || |||   ||||| ||

Db   1212 TGCTCCCAATCCAAGAAGCTATTGATCGACATTTTGAGAGACGAGTGGAAATGGGACGGT 1271

Qy    661 CTAATCATGAGCGACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAGGCCTC 720
          |  | |||  |||||||| ||| ||||| |  ||  | | |||  |   |   ||  |

Db   1272 ATGTTAATGTCCGACTGGTTCGGTACATATACGACTGCCGCAGCTATCAAGAATGGGTTG 1331

Qy    721 GACCTCGAGATGCCCGGACCTCCACGCTTCCGAGGAGAAACACTCAAGTTCAACGT---- 776
          ||  ||||| | || |||||  || | |   ||  |   |  |   ||   ||

Db   1332 GATATCGAGTTCCCTGGACCAACAAGATGGAGAACACGTGCTTTAGTGTCTCACTCTCTC 1391

Qy    777 --CTCCAACGGAAAGCCCTTTATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCAG 834
            |||||  | | |   |  ||   | |   ||||  |  | | |||  |||| ||  |

Db   1392 AACTCCAGAGAACAAATCACTACTGAAGATGTTGATGATCGTGTTAGACAAGTGCTAAAA 1451

Qy    835 TTCGTCAAGAAGTGTGCTGCCTCCGGAGTGAC-----------GGAGAACGGCCCCGAG 882
          |  | |||      || || |     || ||             |||||| ||||| ||

Db   1452 ATGATTAAGTTCGTTGTTGACAATTTAGAGAAAACAGGTATTGTGGAGAATGGCCCAGAA 1511

Qy    883 ACGACTGTCAACAACACCCCCGAAACGGCAGCTCTCCTCCGGAAGGTTGGCAACGAGGGC 942
          | |||   |||||||||   |||||  | |  ||  |  |  || |||     ||

Db   1512 TCTACTTCAAACAACACCAAGGAAACCTCGGACCTGTTGAGAGAGATTGCTGCTGACTCT 1571

Qy    943 ATCGTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCAAGAAGAAGAAGACGCTG 1002
          || ||  | ||||||| | |||||  |||  |||  |  | |  |   |    |

Db   1572 ATTGTTTTATTGAAGAACAAAAACAATTATCTTACCTCTAAAGAAAGAAGACAATATCAT 1631

Qy   1003 ATTGTCGGCCCCAACGCCAAGCAGGCCACATACCACGGCGGAGGCTCTGCCGCACTCAGG 1062
          |  | |||||  || || ||        ||       |||  || || || || ||  | | |

Db   1632 GTCATTGGCCCAAATGCTAAAGCAAAGACTAGTTCCGGTGGTGGTTCAGCATCTATGAAC 1691

Qy   1063 GCCTACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGCTCGAGACGCCGCCATCG 1122
          |||||||  |  ||  |||| | ||| ||  ||  |||  ||||  |  |    |

Db   1692 TCCTACTATGTTGTTTCTCCGTATGAAGGTATCGTCAATAAGCTGGGCAAAGAGGTCGAT 1751
```

```
Qy   1123 TACACCGTCGGCGCCTA--CACCACCGTTCCTCCCATTCTAGGCGAGCAGTGCCTCACGC 1180
Db   1752 TACACCGTAGGGGCCTATTCACACAAATCGATTGGAGGTTTGGCAGAGAGTAGTTTGATC 1811

Qy   1181 CCGACGGCGCTCCGGGCATGCGCTGGAGGGTCTTCAACGA--GCCCCCTGGTACCCCTAA 1238
Db   1812 GATGCTGCAAAACCAGCAGATGCTGAAAATGCTGGATTAATTGCCAAGTTTACTCCAAT 1871

Qy   1239 CCGCCAGCACA----------------------TTGACGAGCTCTTCTTCACCAAGACG 1275
Db   1872 CCAGTAGAAGAGAGATCTGAAGATGAAGAACCATTCCACGTTACCAAAGTCAATAGATCC 1931

Qy   1276 GACATGCACCTGGTGGACTACTACCACCCCAAGGCGGCAGACACG---------TGGTAC 1326
Db   1932 AATGTTCACTTATTTGATTTCAAACATGAGAAAGTGGATCCAAAGAACCCTTACTTTTTT 1991

Qy   1327 GCCGACATGGAGGGCACGTACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTC 1386
Db   1992 GTAACCTTAACCGGACAGTACGTGCCTCAAGAAGATGGTGATTATATCTTCAGTCTTCAA 2051

Qy   1387 GTCTGCGGCACGGCAAAGGCGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAG 1446
Db   2052 GTTTATGGTTCTGGTTTGTTCTACTTAAACGATGAGTTGATTATTGACCAAAAGCACAAC 2111

Qy   1447 CAGGTCCCCGGCGATGCCTTCTTCGGCTCCGCCACCGCGAGGAGACGGGCCGCATCAAT 1506
Db   2112 CAAGAAAGGGGTAGTTTCTGCTTTGGAGCTGGTACCAAAGAAAGAACCAAAAAGTTGACT 2171

Qy   1507 CTCGTCAAGGGCAACACGTACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACC 1566
Db   2172 TTGAAGAAGGGCCAAGTTTATAATGTAAGAGTTGAGTACGGTTCTGGCCCAACTTCAGGT 2231

Qy   1567 CTCAAGGGCGACACCATCGTCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTC 1626
Db   2232 TTGGTTGGGGA---------ATTCGGTGCAGGTGGATTCCAAGCTGGTGTCATTAAGGCG 2282

Qy   1627 ATTGACGACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTC 1686
Db   2283 ATCGATGATGACGAGGAGATTAGAAACGCAGCAGAATTAGCAGCTAAGCATGATAAGGCT 2342

Qy   1687 ATCATCTGCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATG 1746
Db   2343 GTGTTGATAATTGGATTAAATGGTGAATGGGAAACCGAAGGTTATGACAGAGAAAACATG 2402

Qy   1747 AAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACC 1806
Db   2403 GATTTGCCAAAAAGAACAAATGAATTAGTTCGTGCTGTTTTGAAAGCAAATCCAAATACT 2462

Qy   1807 GTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCC 1866
Db   2463 GTTATCGTTAACCAATCAGGTACCCCAGTTGAGTTCCCTTGGTTAGAAGAGGCAAATGCG 2522

Qy   1867 GTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTT 1926
Db   2523 CTAGTTCAAGCTTGGTACGGTGGTAATGAATTGGGTAATGCTATCGCAGATGTCTTGTAC 2582

Qy   1927 GGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAAC 1986
Db   2583 GGTGACGTGGTTCCAAATGGTAAGTTATCGCTCTCTTGGCCATTTAAGTTGCAAGATAAT 2642

Qy   1987 CCCGCGTTTCTCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTGTAC 2046
Db   2643 CCAGCCTTTTTAAACTTCAAGACCGAGTTCGGAAGAGTTGTTTACGGTGAGGATATCTTT 2702

Qy   2047 GTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGC 2106
Db   2703 GTTGGTTATAGGTACTACGAAAAGCTTCAAAGAAAGGTAGCCTTCCCCTTCGGATATGGT 2762

Qy   2107 CTGTCCTACACCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGACGGCAAGCTG 2166
Db   2763 CTATCGTATACAACATTCGAACTAGATATTTCTGACTTCAAGGTAACTGATGATAAGATA 2822

Qy   2167 AGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTC 2226
Db   2823 GATATTTCAGTTGATGTGAAGAATACTGGTGATAAATTTGCTGGCTCCGAGGTGGTGCAA 2882

Qy   2227 TACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTC 2286
Db   2883 GTCTACTTCAGCGCTCTAAACTCTAAGGTCTCGAGACCGGTTAAGGAGTTCAAGGGATTC 2942

Qy   2287 GCAAAGGTCGAACTGCAGCCCGGCGAGACGAAGGCGGTGACAATCGAGGAGCAGGAGAAG 2346
Db   2943 GAAAAAGTTCATTTGGAACCAGGTGAGAAGAAGACAGTTAATATTGAACTAGAATTGAAA 3002

Qy   2347 TACGTCGCTGCGTATTTTGATGAGGAGCGGGATCAGTGGTGTGTCGAAAAGGGTGACTAT 2406
Db   3003 GATGCAATTTCCTACTTTAACGAAGAGCTCGGTAAATGGCACGTTGAAGCAGGTGAATAC 3062

Qy   2407 GAGGTTATCGT 2417
Db   3063 TTGGTTTCAGT 3073
```

RESULT 2
AF486581

```
LOCUS       AF486581                38269 bp    DNA     linear   BCT 21-MAR-2002
DEFINITION  Uncultured bacterium aminoglycoside 3'-N-acetyltransferase gene,
            complete cds.
ACCESSION   AF486581
VERSION     AF486581.1  GI:19569565
KEYWORDS    .
SOURCE      uncultured bacterium.
  ORGANISM  uncultured bacterium
            Bacteria; environmental samples.
REFERENCE   1  (bases 1 to 38269)
  AUTHORS   Cappellano,C.M., Courtois,S., Ball,M., Francou,F.-X., Normand,P.,
            Heylinck,G., Guerineau,M., Martinez,A., Hopke,J., Kolvek,S.,
            August,P.R., Osburne,M.S., Jeannin,P., Simonet,P. and
            Pernodet,J.-L.
  TITLE     Recombinant Environmental Libraries Provide Access to Microbial
            Diversity for Natural Products Drug Discovery
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 38269)
  AUTHORS   August,P.R. and Martinez,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Molecular Biodiversity, Aventis Cambridge
            Genomics Center, 26 Landsdowne Street, Cambridge, MA 02139, USA
FEATURES             Location/Qualifiers
     source          1..38269
                     /organism="uncultured bacterium"
                     /db_xref="taxon:77133"
                     /note="isolated from soil"
     CDS             complement(30829..31617)
                     /note="confers resistance to kanamycin in Escherichia
                     coli; similar to known cognate proteins"
                     /codon_start=1
                     /transl_table=11
                     /product="aminoglycoside 3'-N-acetyltransferase"
                     /protein_id="AAL92107.1"
                     /db_xref="GI:19569566"
                     /translation="MTSRVATRSSLADDLSAIGLADGDAVLVHAALRQVGKIVGGPDA
                     IIDALRDVIGPAGTILGYCDWQLEDELRDDPSMRPHIAAFDPERSRSTRDNGYWPEAL
                     RTTPGALRSGSPGASMAALGGEAEWFTADHALDYGYGPQSPLGKLVEAEGKVLMLGAP
                     LDTMTLLHHAEHLADFPNKRILRYEAPILVDGETVWRWFEEFDTSEPPDGLPEDYFAT
                     IVEAFLATGRGKRGEVGEASSVLVPAAAMVAFGVDWLERWGKTL"
BASE COUNT     6318 a  12117 c  12998 g   6835 t      1 others
ORIGIN

  Query Match             16.2%;  Score 406.2;  DB 1;  Length 38269;
  Best Local Similarity   50.8%;  Pred. No. 2.7e-40;
  Matches 1213;  Conservative    0;  Mismatches 1098;  Indels   78;  Gaps    7;

Qy        17 TTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGATCTGCTGGCTGGTATCG 76
Db     32244 TCGAGGAACTGGTCGACCGGATGACGCTCGAGGAGCAGGTGAGCCTGTTATCGGGCGAGG 32303
```


Qy 77 ACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCTCTCCGCTTTACAGATG 136
Db 32304 ATTTCTGGTCGCTGCCGGCGATCGATCGGCTCGGCGTCGGCAAGCTCCGCGTCACCGACG 32363
Qy 137 GCCCCAACGGCGTAAGAGGGACCAAGTTCTTCAATGGC---GTCCCTGCGGCCTGCTTCC 193
Db 32364 GGCCGAACGGGGCGCGTGGCGGCGGCTCGCTGATCGGCGGGGTCAAGTCGGCGGCGTTCC 32423
Qy 194 CTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGA 253
Db 32424 CGGTGGGGATCGCGCTCGGCGCCAGCTGGAACGTCCGGCTGCTTACCGAGGTCGGCACGG 32483
Qy 254 TGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACA 313
Db 32484 CTTTGGCCGAGGAGGTGAAATCGAAGGGGGCGCATGTGAGCCTCGCGCCGACGGTGAACC 32543
Qy 314 TGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGG 373
Db 32544 TCCATCGTTCGGTCAGCAACGGGCGCAATTTCGAGTGTTATTCGGAGGACCCGATCCTCA 32603
Qy 374 CGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGA 433
Db 32604 CCGCCAGCCTGGCGGTCGGCTATATCGAGGGGTTGCAGGGGCAGGGCGTCTCGGCGACGA 32663
Qy 434 TCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCA 493
Db 32664 TCAAGCACTTTGCCGGCAATGAATCCGAGATCGAGCGAACCACCATCAGCTCCGAGATCG 32723
Qy 494 CGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCC 553
Db 32724 ACGAGCGGGCGCTGCGTGAGCTCTACCTCGTGCCGTTCGAGTGGGCGGTGAAGCGGGCCG 32783
Qy 554 AGCCGGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACC 613
Db 32784 GCACCTGGGGCATCATGAGCTCCTACAACAAGCTCAATGGCACCTTCACCAGCGAGAACC 32843
Qy 614 CTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCG 673
Db 32844 CCTGGCTGCTCGACCAGGTGCTGCGCAGGGAATGGCACTATGACGGCGTGGTGATGTCGG 32903
Qy 674 ACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGC 733
Db 32904 ACTGGTTCGGCTCGCACTCGACCGCGCCGACCGTCAATGCCGGGCTCGACCTCGAAATGC 32963
Qy 734 CCGGACCTCCACGCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCT 793
Db 32964 CCGGGCCGCCGCGCGACCGTGGCGGGAAGCTGGTGGCGGCGGTGGAGGCCGGCGAGGTCA 33023
Qy 794 TTATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCTG 853
Db 33024 GCCGCGAAACGCTGCGGACCCGGGTGCTCGACGTGCTGCGGCTGATTGAGCGGACCGGGG 33083
Qy 854 CCTCCGGAGTGACGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCGAAACGGCAG 913
Db 33084 CGCTGGACG---ATCACAGCGAGTGGGTGGAGCGGGCCGACGATCGCGCCGAGCACCGGG 33140
Qy 914 CTCTCCTCCGGAAGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGAACAACGTTC 973
Db 33141 CGCTGATCCGTCGCGCCGGCGCGGAAGGCGCGGTGCTGCTGAAAAATGCGGGCGAACTGC 33200
Qy 974 TGCCCTTGAGCAAGAAGAAGAAGACGCTGATTGTCGGCCCCAACGCCAAGCAGGCCACAT 1033
Db 33201 TGCCATTGAGCGGGTCGGGCAAGATCGCGGTGATCGGGCCGAACGCTAAAGTGGCGCAGA 33260
Qy 1034 ACCACGGCGGAGGCTCTGCCGCACTCAGGGCCTACTACGCAGTCACTCCCTTTGACGGCC 1093
Db 33261 TCATGGGCGGCGGCAGCGCGCAGCTCAACCCGCACTATCGGGTTTCGCCGTGGGATGGTC 33320
Qy 1094 TCAGCAAGCAGCTCGAGACGCCGCCATCGTACACCGTCGGCGCCTACACCACCGTTCCTC 1153
Db 33321 T----------GGTCGCGGCCTTGGGCGAAGAACGCCTCAGCTATGCGG 33359


Qy 1154 CCATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGGAGGGTCT 1213
Db 33360 CGGGTGCGACCAACCACCGCTTCGAGCCGCTGCTGCGCGGCGGCTTCACGGCTGAGTATT 33419
Qy 1214 TCAACGAGCCCCCTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGA 1273
Db 33420 TCGCCAATGAGACTCTCGCCGGCACGCCGGTGTTCAGCGAGACGCTGGAAGAGGCGCAGG 33479
Qy 1274 CGGACATGCACCTGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTACGCCGACA 1333
Db 33480 CGTTCTGGATCGGGCGAGTCGGTGGCGGCAAGGTCGATCCACTGCATTTCTCGGCCAGGC 33539
Qy 1334 TGGAGGGCACGTACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTCGTCTGCG 1393
Db 33540 TCACCGGCCGGTTCACCCCGGAGGCGAGCGGCGAACATCGCGTCGGCATCTACAGCGCCG 33599
Qy 1394 GCACGGCAAAGGCGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCC 1453
Db 33600 GTTTTTCGCGGGTGCTGGTCGACGGCAAGCTGGTGGCCGATGCGTGGACGAACTGGCGCA 33659
Qy 1454 CCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAATCTCGTCA 1513
Db 33660 AAGGCCGCACCTTCTTTGAGGAAGGCTGCGACGAGGTCGTCGGCACGATCGATCTCGAAG 33719
Qy 1514 AGGGCAACACGTACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACCCTCAAGG 1573
Db 33720 CCGGCCGCAGCTACGAAGTGGTGATCGAGTTCGCCACCAAGCCGTTCGCGACGCTGGGGC 33779
Qy 1574 GCGACACCATCGTCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACG 1633
Db 33780 TCGCGGCCTTCGCCGCCGGCATCGGCCTGCCGCTTGGCGAC---------- 33820
Qy 1634 ACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCT 1693
Db 33821 -----GCGGCGATCGCCGAAGCGGTCGCAGCGGCGCGTGACGCCGAGGTGGCGGTGGTAT 33875
Qy 1694 GCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCC 1753
Db 33876 TCGTCGGGCGCAACGGCGAGTGGGACACCGAGGGCAGCGACCTCGATTCGATCGTCTTGC 33935
Qy 1754 CCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCG 1813
Db 33936 CCGGCCGGCAGGACGAGCTGATCGCCGCGGTTGCGGTGGTGAACCCGCGCACCGTGGTGG 33995
Qy 1814 TCATGCAGACGGGCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCC 1873
Db 33996 TGCTGCAGACCGGCGGGCCGGTGGAAATGCCGTGGCTCGGTGACGTCGCCGCGGTGCTCG 34055
Qy 1874 AGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACT 1933
Db 34056 AGGCCTGGTATCCGGGGCAGGAGGCGGGCAACGCCATCGCTGACGTGCTGACCGGCGCGG 34115
Qy 1934 ACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAACCC----- 1988
Db 34116 CCGAGCCGGGCGGCCGGCTGCCGCAGAGCTTTCCGGTCAAGTGGGCGGACAATCCGACCA 34175
Qy 1989 -------CGCGTTTCTCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACG 2041
Db 34176 ACTCGCAGGATCGCGAAATCTATCCCGGCCTCGACGGCAAGGTGCGCTACGAGGAGGGGG 34235
Qy 2042 TCTACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCC 2101
Db 34236 TGTTCATCGGCTATCGCCACTACGATCGCCATGGCATCGCCCCGCTGTTCCCGTTCGGCT 34295
Qy 2102 ACGGCCTGTCCTACACCACTTTTGCCTTTTCCAATCT----------CTCCGTGTCTC 2149
Db 34296 TCGGACTTGGCTACACCAGCTTTGCGGTGTCGGAGCTTGAACTCGACGCCTCGGGCTTCG 34355
Qy 2150 ACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCG 2209
Db 34356 AAGCGGATGGCAGTGTGAGCGTCTCGGTCACGGTCGAGAATACCGGCCGGCGCGCCGGCT 34415
Qy 2210 CACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCA 2269

```
Db  34416 CGGAAGTGGTGCAGGTCTACGTGTCGGCGCCGGATGCG---ACGGTGTCGCGGCCGGCCA 34472

Qy   2270 AGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCAGCCCGGCGAGACGAAGGCGGTGACAA 2329
Db  34473 AGGAGCTGAAGGCTTTCGAGAAGCTGCATCTGGCGCCGGGCGAGCGGCAGCGGGTGACGC 34532

Qy   2330 TCGAGGAGCAGGAGAAGTACGTCGCTGCGTATTTTGATGAGGAGCGGGA 2378
Db  34533 TGGAGCTCGATGACCGGGCTTTCGCTTACTACCGCACCGAAGCGCAGCA 34581


RESULT 3
SPU33009
LOCUS       SPU33009               39305 bp    DNA     linear   PLN 14-NOV-1995
DEFINITION  Schizosaccharomyces pombe cosmid 1683.
ACCESSION   U33009
VERSION     U33009.1  GI:1063616
KEYWORDS    .
SOURCE      Schizosaccharomyces pombe.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 39305)
  AUTHORS   McCombie,W.R.
  TITLE     Sequence analysis of a region of the fission yeast genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 39305)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1995) W. Richard McCombie, Cold Spring Harbor
            Lab, 1 Bungtown Rd, P.O. Box 100, Cold Spring Harbor, NY 11754, USA
FEATURES             Location/Qualifiers
     source          1. .39305
                     /organism="Schizosaccharomyces pombe"
                     /db_xref="taxon:4896"
                     /chromosome="2"
                     /clone="cosmid 1683"
                     /clone_lib="Mizukami, et al., 1993. Cell 73:121-132."
BASE COUNT    12234 a   7052 c   7011 g  13008 t
ORIGIN

  Query Match           15.7%;  Score 393.6;  DB 8;  Length 39305;
  Best Local Similarity  49.0%;  Pred. No. 8.8e-39;
  Matches 1199;  Conservative    0;  Mismatches 1229;  Indels   18;  Gaps    5;

Qy       1 ATGGCTGATATTGATGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGAT 60
Db    8497 ATGATGGAACATGATGTTGAAGATTTGATCAATCAATTGGATATAAGTGAGAAGGCTATG 8556

Qy      61 CTGCTGGCTGGTATCGACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCT 120
Db    8557 CTGCTTTCCGGCACTGACCTTTGGCATACTGCAGCAATTCCTCGTCTTAACATTCCTTCC 8616

Qy     121 CTCCGCTTTACAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCTTCAATGGCGTCCCT 180
Db    8617 ATTAGATTATCAGACGGCCCTAATGGTATCCGTGGAACTAGCTTCTTCAATTCTTCACCA 8676

Qy     181 GCGGCCTGCTTCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAA 240
Db    8677 TCCGCTTGTTTTCCCTGTGGGACAGCGCTAGGGGCTACTTTCGACAAAAAGTTACTATTC 8736

Qy     241 GAGGCAGGTAAGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGC 300
Db    8737 GAAGTTGGTGAATATTTAGCAGAAGAAGCAAAAGCGAAAGGTGTTAGTGTGGTTTTGGGT 8796

Qy     301 CCGACTATCAACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAG 360
Db    8797 CCAACGGTAAATATCCATCGTGGACCCTTAAATGGTAGGGGTTTTGAATCATTTTCTGAA 8856

Qy     361 GATCCGTTCCTGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGA 420
Db    8857 GATTCTACTTTATCAGGTCTTGCAGCTAGTTATGTTATTCTTGGATTGCAAAGTAAAAAC 8916

Qy     421 GTGCAGGCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATGATGGTG 480
Db    8917 GTCCAAGCATGCATTAAGCATTTTGTGTGTAATGATATGGAAGATGAAAGGAACTCTGTC 8976

Qy     481 CAGAGCATCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATTGCT 540
Db    8977 AGTATTGACGTCTCACAAAGGGCACTTAGAGAAGTATATCTTATGCCATTTCAGTTGGCA 9036

Qy     541 GTGCGAGACTCCCAGCCGGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCGTGTCG 600
Db    9037 TGCAAATATTCAAATTTCAAATCACTTATGACTTCTTACAACAAGGTGAATGGTGAACAC 9096

Qy     601 TGCAGCGAGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGC 660
Db    9097 GTTTCCCAATCTCGAATTTTATTAGACAATATTTTAAGAAAAGAGTGGGAGTGGAAAGGT 9156

Qy     661 CTAATCATGAGCGACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAGGCCTC 720
Db    9157 ACTATAATATCTGATTGGTTTGGAACTTATTCGTTGAAGAAAGCTATAGATGCAGGTTTG 9216

Qy     721 GACCTCGAGATGCCCGGACCTCCACGCTTCCGAGGAG---AAACACTCAAGTTCAACGTC 777
Db    9217 GATTTGGAGATGCCAGGAAAGCCGAGATTTCGAAATGTCAACACTATTCAACACTTAGTA 9276

Qy     778 TCCAACGGAAAGCCCTTTATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCAGTTC 837
Db    9277 GGATCCAAAGAGCTTTCGGAGTCAATTTTAGATGAAAGAGCAAAGAATGTCCTGAAGCTT 9336

Qy     838 GTCAAGAAGTGTGCTGCCTCCGGAGTGACGGAGAACGGCCCCGAGACGACTGTCAACAAC 897
Db    9337 GTGAAACATTCCTGGCAGAATACAGAAGCAGAAAATCATTGTGAGCTAAATAATGATTCC 9396

Qy     898 ACCCCCGAAACGGCAGCTCTCCTCCGGAAGGTTGGCAACGAGGGCATCGTGCTGCTGAAG 957
Db    9397 TCTTGTTTAAGAGAAGC---ACTTAAAAAATTTGCAAGCCAGTCCATTGTTCTCTTGAAA 9453

Qy     958 AACGAGAACAACGTTCTGCCCTTGAGCAAGAAGAAGAAGACGCTGATTGTCGGCCCCAAC 1017
Db    9454 AATAAAAAGAAACTTTTGCCATTGTCAAGGAAGGGCACGTTTGCAGTTATTGGACCTAAC 9513

Qy    1018 GCCAAGCAGGCCACATACCACGGCGGAGGCTCTGCCGCACTCAGGGCCTACTACGCAGTC 1077
Db    9514 GCAAAGGTTTGTAATTACAGCGGAGGTGGATCGGCAAATTTGAAGCCTTATTATACTGTG 9573

Qy    1078 ACTCCCTTTGACGGCCTCAGCAAGCAGCTCGAGACGCCGCCATCGTACACCGTCGGCGCC 1137
Db    9574 AGTATGTATGATGGCATAGCAGCGAAAATAGACGGTGTGCCAGAGTATGCTCTTGGTTGC 9633

Qy    1138 TACACC------ACCGTTCCTCCCATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCT 1191
Db    9634 CATAACTATCTAAACCTTCCAAATATTGCAAATTTATTAGTCAACCCTAGAACCGGCAAG 9693

Qy    1192 CCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCCCTGGTACCCCTAACCGCCAGCACATT 1251
Db    9694 CATGGATACGTTGCGAAGTTTTATCTTGAACCTGCAACCTCCGAAAATAGAACTTTGATA 9753

Qy    1252 GACGAGCTCTTCTTCACCAAGACGGACATGCACCTGGTGGACTACTACCACCCCAA---G 1308
Db    9754 GACGATTACGATTTAGAAGATGGAGTTGTTCGTTTTTACGATTATTGTAACGACAAAATG 9813

Qy    1309 GCGGCAGACACGTGGTACGCCGACATGGAGGGCACGTACACCGCCGACGAGGACTGCACC 1368
Db    9814 AAGGATGGCTATTTCTATATAGACATTGAAGGTTATTTAATTCCTGACGAGGACGCTGTT 9873

Qy    1369 TACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGGCGTACGTAGACGACCAGCTCGTC 1428
Db    9874 TATGAATTTGGGATATCTGTGTTTGGAACGGCTCTGCTGTTTATTGATGATGTCTTATTA 9933

Qy    1429 GTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAG 1488
Db    9934 ATCGACAACAAGACAAAGCAAACCCCTACGAATCATACTTTTGAATTTGGAACAATAGAA 9993
```

```
Qy   1489 GAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGTACAAGTTCAAGATCGAGTTCGGC 1548
          ||||        ||  || |    |||||  |  | ||| ||  |   | ||||||  ||
Db   9994 GAGAGAAATTCTATTTATTTAAAGAAGGGGAGAAAGTATAATGTACGGGTCGAGTATGGA 10053

Qy   1549 TCCGCACCCACCTACACCCTCAAGGGCGACACCATCGTCCCCGGCCACGGCTCCCTCCGC 1608
             ||| |||| ||||| ||         |          |          ||       |
Db  10054 AGTGCAGCCACTTACACGCTGTCAACTAATCTGTCTCCCAGTACGGGAGGAAGATACTCT 10113

Qy   1609 GTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCC 1668
           | ||| |    || || |||||     |  | || || ||  |  | ||     | |||
Db  10114 ATTGGCTGTGTTAAAGTTATTGATCCTGAAACGGAGATAGATTACGCAGTGAGGGTTGCC 10173

Qy   1669 AAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGC 1728
           ||       ||     || ||  | || | |||  | |  || || ||||| || || ||
Db  10174 AAATCAGTTGATTGTGTGATTTTATGTGTGGGTTTGACAGCAGAATGGGAAACGGAAGGT 10233

Qy   1729 GCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCC 1788
           |  ||  |    |  ||||   | ||     |    ||| | ||  ||       |
Db  10234 GAAGATAGAAAAACAATGACTTTACCATCGTTATCTGACAAACTGGTTTATTCTATTCTG 10293

Qy   1789 GCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATGCCCTGG 1848
             | || || || || || || || || |  ||  | ||||| ||    |||||||| |||
Db  10294 CAATCCAATCCCAATACAGTAGTTGTGACTCAATCAGGCACTCCAATTGAGATGCCATGG 10353

Qy   1849 CTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCC 1908
           ||   |   | | | || |  | ||    ||||||  ||| |||||    || ||  |
Db  10354 ATCTCTGAAGCCCACACCTTATTGCACATTTGGTACAACGGTAACGAATTAGGAAATGCT 10413

Qy   1909 ATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCC 1968
           | ||  || |  | ||||| ||  | ||||| |  || ||  || |  | |  |||||
Db  10414 TTAGCAAACATTATTTTTGGAGAGCAAAACCCTTGTGGAAAATTGCCAATAACATTCCCT 10473

Qy   1969 AAGCGCCTGCAGGACAACCCCGCGTTTCTCAACTTCCGCACCGAGGCCGGGCGCACGCTG 2028
           ||    ||| |||| || || || | |||    || ||  |        || |      |
Db  10474 AAAAAACTGAAGGATAATCCTGCCTATCTTTCATTTCGTTCAAGCCGTGGCCATTGTGTT 10533

Qy   2029 TACGGCGAGGACGTCTACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAAT 2088
           || || || ||||||  || ||||| | ||| || ||    |  || | ||| |||
Db  10534 TATGGAGAAGACGTCTTTGTTGGGTATAAGTATTATGAAGCCGTAGAAAGGGAGGTCTTG 10593

Qy   2089 TTCCCCTTTGGCCACGGCCTGTCCTACACCACTTTTGCCTTTTCCAATCTCTCCGTGTCT 2148
           || || |||||  | || || || || || |||||||  || |||||||  |    ||
Db  10594 TTTCCATTTGGATATGGGCTTTCATATACTACTTTTGAATTATCCAATCTTTATTTGAAG 10653

Qy   2149 CACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGC 2208
           |    |  | | | | ||  |      |      |   ||| || ||| || || | ||
Db  10654 AATTGTGGTGAAAGGTTAAGGATCGATTTGGAGATTTCGAATACTGGCCCCATGTCTGGT 10713

Qy   2209 GCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTC 2268
           |||  |   |    ||  | ||  |    |   ||   |          |||| || || ||
Db  10714 GCAGAAATTATACAAGTATATATTTCTCAAATC---GTTCGTTCAGTTAATCGACCTGTT 10770

Qy   2269 AAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCAGCCCGGCGAGACGAAGGCGGTGACA 2328
           || || || | |  | || | || || |    |    ||    ||||| ||     | | |
Db  10771 AAAGAACTAAAAGAGTTTTCTAAAGTGGTTTTATGTCCTAAAGAGACCAAATTAATTAGA 10830

Qy   2329 ATCGAGGAGCAGGAGAAGTACGTCGCTGCGTATTTTGATGAGGAGCGGGATCAGTGGTGT 2388
           || |||  | |    ||||| |   || | | || ||||||       ||  | ||||||
Db  10831 ATTGAGTTGGATATAAAGTATGCTACTTCTTTTTATGATGAACTTAATGAAAAATGGTGT 10890

Qy   2389 GTCGAAAAGGGTGACTATGAGGTTATCGTGAGCGACAGCAGCGCAG 2434
             ||| |||| || ||  | ||| | ||  |   ||||||  ||
Db  10891 TCTGAAGAGGGCGAGTACAACGTTCTTGTCGGAACCAGCAGTAAAG 10936
```

RESULT 4
SPBC1683
LOCUS SPBC1683 41799 bp DNA linear PLN 23-AUG-2000
DEFINITION Schizosaccharomyces pombe cosmids c1683.
ACCESSION AL355920
VERSION AL355920.1 GI:7801293
KEYWORDS adenosine deaminase; alpha-glucosidase; beta-glucosidase precursor (EC 3.2.1.21); ferric reductase transmembrane component; frp1; ght4; glucose transporter protein; glycosyl hydrolase family 3; inosine-uridine preferring nucleoside hydrolase; isocitrate lyase (EC 4.1.3.1); MFS allantoate permease; MFS inorganic phosphate transporter; MFS transporter of unkown specificity; NCS1 allantoate transporter; TF1-type LTR; transcriptional activator; zinc finger.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE 1 (bases 1 to 41799)
AUTHORS McCombie,W.R.
TITLE Sequence analysis of a region of the fission yeast genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 41799)
AUTHORS McCombie,W.R. and Lyne,M.H.
TITLE Direct Submission
JOURNAL Submitted (11-APR-2000) W. Richard McCombie, Cold Spring Harbor Lab, 1 Bungtown Rd, P.O. Box 100, Cold Spring Harbor, NY 11754, USA and Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA
COMMENT Notes:
Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
CDS are numbered using the following system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c (complementary strand).
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid SPBC1683 is overlapped at its 5' end by cosmid SPBC359 and at its 3' end by cosmid SPBC1198.

FEATURES Location/Qualifiers

source 1. .41799
/organism="Schizosaccharomyces pombe"
/db_xref="taxon:4896"
/chromosome="II"
/map="IIL"
/clone="cosmids 359 (partial), 1198 and 1683"
/clone_lib="Mizukami, et al., 1993. Cell 73:121-132."

misc_feature 1. .500
/note="nominal overlap with cosmid SPBC359 S. pombe chromosome II"

mRNA 1351. .1543
/note="mRNA from spc08587 1 193"

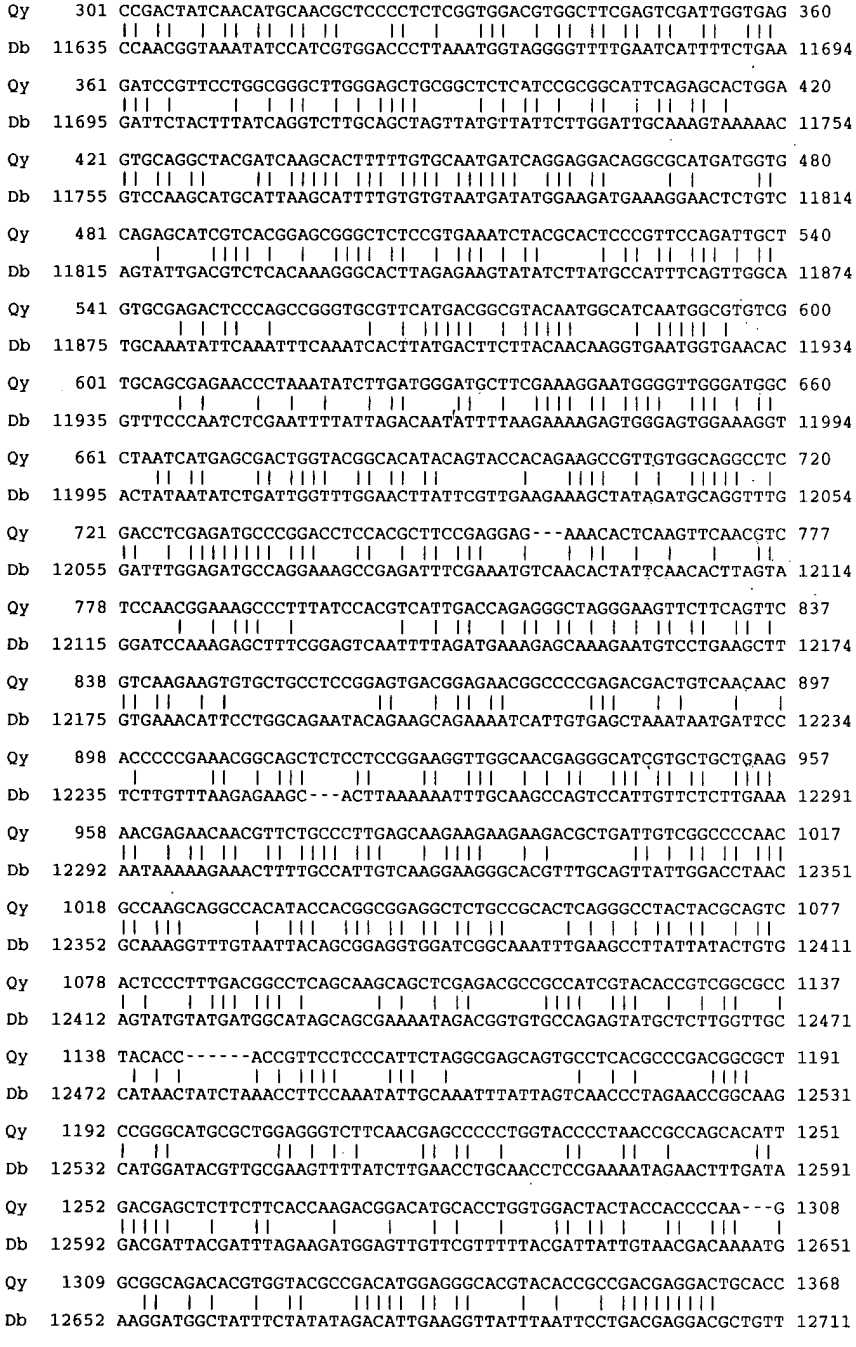
gene 3226. .4947
/gene="SPBC1683.01"

CDS 3226. .4947
/gene="SPBC1683.01"
/note="SPBC1683.01, len:573, SIMILARITY:Schizosaccharomyces pombe, YAEC_SCHPO, putative inorganic phosphate transporter c23d3.12., (559 aa), fasta scores: opt: 2395, E():0, (70.0% identity in 574 aa)"
/codon_start=1
/label=SPBC1683.01
/product="MFS inorganic phosphate transporter"
/protein_id="CAB91163.1"
/db_xref="GI:7801294"
/translation="MKLNIFKSHGDSNTAEERPVPLEQVEAEDQQHENRFWLGLTAKE
FRLMMLAGVGFFLDSYDLFIINLVTPIFEYLYWGGIEKGPNGKGHYPSGIRGLVNAAS
NIGNIFGQLMFGFMGDFFGRKFVYGKEMIIVIIATILLIAMPKSIHSPLSKMMWVFCW
RWLLGVGIGGDYPMSAAITSERSKLNRRGTLISLIFAFQGFGTLAGAIVTIILLGCFE
HPLNREGHYRKLEGVWRLQFGLALVPAIGVLIPRLMMEETQKFKNSQQLNSGDNRDPK
TSLNFEDDELVKNPSVTKGHPEIHESSENYLSRSNTVENEPENIEKQFESVSAPANRS
GFIQYFRQWHHFKHLLGTSVCWFLLDIAFYGVNLNQSVILKNIGFSSGTNEYRTLMKN
AIGNLIIAVAGYVPGYWFNVFLVEILGRKWIQLQGFVITGLMFAILAGRWNEISTGGR
FACFVIAQLFSNFGPNSTTFIYPAEVFPARVRGTAHGISAALGKCGAILASLLFNFLT
SIIGYGNVMWIFCGCMWGGILFTLLLPETKGRDADEIDRVELFYGGDGKVECNSKWKS
WYVNGIF"

misc_feature join(3364. .4004,4134. .4872)
/gene="SPBC1683.01"
/note="Match to PF00083 sugar_tr, Sugar (and other) transporter Score 132.74"
LTR complement(5308. .5478)
/note="partial degraded TF1-type LTR"
gene 5967. .6986
/gene="SPBC1683.02"
CDS 5967. .6986
/gene="SPBC1683.02"
/note="SPBC1683.02, len:339, SIMILARITY:Saccharomyces cerevisiae, ADA_YEAST, probable adenosine deaminase, (347 aa), fasta scores: opt: 723, E():0, (37.6% identity in 351 aa)"
/codon_start=1
/label=SPBC1683.02
/product="putative adenosine deaminase"
/protein_id="CAB91164.1"
/db_xref="GI:7801295"
/translation="MTDIERFIEKLPKAELHLHLEGTLEAELKLKLSHRNKIPLKQSSIEEIKESYNFHDLASFLEVYYEGVELLLHEQDFYDLCYQYLRKAASQNVVYAEMFFDPQLHTRRGISFETVIKGLIRARDDAMRDFHIYSQLIMCFIREMSFENAEETLNASLPYKSEIIGIGLDSNEENNPPIKFLKVFQRARQLGYRLTCHCDLHQKNTTTHIRQALEDIGVERIDHGINILDDPELIKLALERNIPFTVCPFSNEIVYPGKAQPEIRIMLDTGLKVTINSDDPAYMHCFYITENFNLAQKGASLTKKELVQICRNSFEAAWISEEKRNHYLEALNSFASAYD"
misc_feature 6141. .6956
/gene="SPBC1683.02"
/note="Match to PF00962 A_deaminase, Adenosine/AMP deaminase Score 164.32"
gene complement(7448. .9007)
/gene="SPBC1683.03c"
CDS complement(7448. .9007)
/gene="SPBC1683.03c"
/note="SPBC1683.03c, len:519, SIMILARITY:Saccharomyces cerevisiae, Q08902, orf yor378w., (515 aa), fasta scores: opt: 2052, E():0, (58.0% identity in 509 aa)"
/codon_start=1
/label=SPBC1683.03c
/product="MFS transporter of unkown specificity"
/protein_id="CAB91165.1"
/db_xref="GI:7801296"
/translation="MMNYVKSLGTYNTRNMEKESSKDLVQNEDTPLPVQKISMSLFHEIVFVFIACTAQLMTQAGLGQSIAPNNIIGKSLGTTNPGQLSWFPASYSLTVGTFILIAGRLGDIYGHKKMFVLGYIWFCIWSLISGFSYYAKSVIMFDVCRALTGIAPAFLLPNALALLGRVYPPGKKKNLIPALFGATAPNGFLLGSVFSGIFAQLSWWPWTYWTTAIVCIVFAIIGYFAIPHIEADEVEEKQKFDYLGAFFGVSGLVLINFSWNQGPVVGWQVPYVYILLIIGFLSLVVFVLVEKRVVQPILAPSMMNSEMGCVLICVAAGWACFGIWMYYLWQFLENLRFATPLLVTAQLTPVGISGCAAALTTGYLLKRLKPTKIMVVSMIAFTVGTILIATAPIHQTYWAQTFVSIIVTPWGMDMSFPAATLMLSDFVPKQHQGIAASLVSTVVNYSISIGLGIAGTVETHLNHKGVDLIRGYRSAWYMGTGFGGLGVCVAILTVFASYLKVGQQKSPKFPLIMKNTKAV"
gene 11335. .13833
/gene="SPBC1683.04"
CDS 11335. .13833
/gene="SPBC1683.04"
/note="SPBC1683.04, len:832, SIMILARITY:Kluyveromyces marxianus, BGLS_KLUMA, beta-glucosidase precursor, (845 aa), fasta scores: opt: 1274, E():0, (46.3% identity in 848 aa)"
/codon_start=1
/label=SPBC1683.04
/product="beta-glucosidase precursor (EC 3.2.1.21); glycosyl hydrolase family 3"
/protein_id="CAB91166.1"
/db_xref="GI:7801297"
/translation="MMEHDVEDLINQLDISEKAMLLSGTDLWHTAAIPRLNIPSIRLSDGPNGIRGTSFFNSSPSACFPCGTALGATFDKKLLFEVGEYLAEEAKAKGVSVVLGPTVNIHRGPLNGRGFESFSEDSTLSGLAASYVILGLQSKNVQACIKHFVCNDMEDERNSVSIDVSQRALREVYLMPFQLACKYSNFKSLMTSYNKVNGEHVSQSRILLDNILRKEWEWKGTIISDWFGTYSLKKAIDAGLDLEMPGKPRFRNVNTIQHLVGSKELSESILDERAKNVLKLVKHSWQNTEAENHCELNNDSSCLREALKKFASQSIVLLKNKKKLLPLSRKGTFAVIGPNAKVCNYSGGGSANLKPYYTVSMYDGIAAKIDGVPEYALGCHNYLNLPNIANLLVNPRTGKHGYVAKFYLEPATSENRTLIDDYDLEDGVVRFYDYCNDKMKDGYFYIDIEGYLIPDEDAVYEFGISVFGTALLFIDDVLLIDNKTKQTPTNHTFEFGTIEERNSIYLKKGRKYNVRVEYGSAATYTLSTNLSPSTGGRYSIGCVKVIDPETEIDYAVRVAKSVDCVILCVGLTAEWETEGEDRKTMTLPSLSDKLVYSILQSNPNTVVVTQSGTPIEMPWISEAHTLLHIWYNGNELGNALANIIFGEQNPCGKLPITFPKKLKDNPAYLSFRSSRGHCVYGEDVFVGYKYYEAVEREVLFPFGYGLSYTTFELSNLYLKNCGERLRIDLEISNTGPMSGAEIIQVYISQIVRSVNRPVKELKEFSKVVLCPKETKLIRIELDIKYATSFYDELNEKWCSEEGEYNVLVGTSSKDIALTGKFTLPKTIHWTGL"
misc_feature 11428. .13776
/gene="SPBC1683.04"
/note="Match to PF00933 Glyco_hydro_3, Glycosyl hydrolase family 3 Score 687.70"
mRNA 16074. .16191
/note="mRNA from spc06472 118 1"
gene 17673. .19352
/gene="SPBC1683.05"
CDS 17673. .19352
/gene="SPBC1683.05"
/note="SPBC1683.05, len:559, SIMILARITY:Saccharomyces cerevisiae, THIY_YEAST, putative thiamine transporter yor192c., (599 aa), fasta scores: opt: 828, E():0, (26.8% identity in 512 aa)"
/codon_start=1
/label=SPBC1683.05
/product="NCS1 allantoate transporter"
/protein_id="CAB91167.1"
/db_xref="GI:7801298"
/translation="MEDPKSDEKFDIGISEKNLDVGFGESSSVDVPVKGRFASFLKKLELSSGPEKENIDLRPTPPDRRHYSALDIIYLWSCNGISASAFRTGTSYMEMGLSPKQALAALIAGNVFIAMPMTLNGLFGSHYHIPFAVQSRASFGYYFNTLIILLRFIAGLFYYGTNVYTGAECVQTILYAIFKSFRSYKNRLPADAGITSDFLISYFVYWVISFPPHLIRPEYLQRFFLIKSISTYIACFAMLIFLLCNVGSHVVWDQPATVSGRSWSWVFMCALNSSVAGFSTLAVNVNDFTRYVKHPKTPYVQMLILPLVAAVSAPIGIVSGVASKIMYGTAMWDPLQIANNWTSRGGRAAAFFMGLTYLVSMIAQNISDNTVAAANDLLYFFPRYLDIRRAQVIVIIIGAWAIVPWKILQNGTAFLAFLGSLSIFLGPAAGIFVADKFKNHHKYDIDEFYNPSGIYRYNKLGLNWRALIAFLCACVPLIPGMAMSINPSITMPDGVIHLYYIGYFYSFMTAFLIYWGLNLVFPAKETLLEEAVYPPKSNAELVDPSTLSGKDKFWYYIDY"
gene complement(20737. .21669)
/gene="SPBC1683.06c"
CDS complement(20737. .21669)
/gene="SPBC1683.06c"
/note="SPBC1683.06c, len:310, SIMILARITY:Escherichia coli., YBEK_ECOLI, hypothetical 33.8 kd protein in leus-gltl intergenic region., (311 aa), fasta scores: opt: 838, E():0, (44.5% identity in 308 aa)"
/codon_start=1
/label=SPBC1683.06c
/product="putative inosine-uridine preferring nucleoside hydrolase"

Query Match 15.7%; Score 393.6; DB 8; Length 41799;
Best Local Similarity 49.0%; Pred. No. 8.6e-39;
Matches 1199; Conservative 0; Mismatches 1229; Indels 18; Gaps 5;

```
Qy      1 ATGGCTGATATTGATGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGAT 60
          |||   ||   |||||||||  |   |  | ||  |  |  |     |    |||||||
Db  11335 ATGATGGAACATGATGTTGAAGATTTGATCAATCAATTGGATATAAGTGAGAAGGCTATG 11394

Qy     61 CTGCTGGCTGGTATCGACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCT 120
          |||||  | || |  ||| | ||||| ||    ||  | ||    | |    | || ||
Db  11395 CTGCTTTCCGGCACTGACCTTTGGCATACTGCAGCAATTCCTCGTCTTAACATTCCTTCC 11454

Qy    121 CTCCGCTTTACAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCTTCAATGGCGTCCCT 180
            |  | ||  |||| ||||| || ||  |  | || || |  |||||||||      ||
Db  11455 ATTAGATTATCAGACGGCCCTAATGGTATCCGTGGAACTAGCTTCTTCAATTCTTCACCA 11514

Qy    181 GCGGCCTGCTTCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAA 240
          | || || || || || || || ||  |||| ||  | || ||| || |||   | ||
Db  11515 TCCGCTTGTTTTCCCTGTGGGACAGCGCTAGGGGCTACTTTCGACAAAAAGTTACTATTC 11574

Qy    241 GAGGCAGGTAAGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGC 300
          || |  ||| |     | |   |||| || |  || ||  |||    |||| |  | ||
Db  11575 GAAGTTGGTGAATATTTAGCAGAAGAAGCAAAAGCGAAAGGTGTTAGTGTGGTTTTGGGT 11634
```


Qy 301 CCGACTATCAACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAG 360
Db 11635 CCAACGGTAAATATCCATCGTGGACCCTTAAATGGTAGGGGTTTTGAATCATTTTCTGAA 11694
Qy 361 GATCCGTTCCTGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGA 420
Db 11695 GATTCTACTTTATCAGGTCTTGCAGCTAGTTATGTTATTCTTGGATTGCAAAGTAAAAAC 11754
Qy 421 GTGCAGGCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATGATGGTG 480
Db 11755 GTCCAAGCATGCATTAAGCATTTTGTGTGTAATGATATGGAAGATGAAAGGAACTCTGTC 11814
Qy 481 CAGAGCATCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATTGCT 540
Db 11815 AGTATTGACGTCTCACAAAGGGCACTTAGAGAAGTATATCTTATGCCATTTCAGTTGGCA 11874
Qy 541 GTGCGAGACTCCCAGCCGGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCGTGTCG 600
Db 11875 TGCAAATATTCAAATTTCAAATCACTTATGACTTCTTACAACAAGGTGAATGGTGAACAC 11934
Qy 601 TGCAGCGAGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGC 660
Db 11935 GTTTCCCAATCTCGAATTTTATTAGACAATATTTTAAGAAAAGAGTGGGAGTGGAAAGGT 11994
Qy 661 CTAATCATGAGCGACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAGGCCTC 720
Db 11995 ACTATAATATCTGATTGGTTTGGAACTTATTCGTTGAAGAAAGCTATAGATGCAGGTTTG 12054
Qy 721 GACCTCGAGATGCCCGGACCTCCACGCTTCCGAGGAG---AAACACTCAAGTTCAACGTC 777
Db 12055 GATTTGGAGATGCCAGGAAAGCCGAGATTTCGAAATGTCAACACTATTCAACACTTAGTA 12114
Qy 778 TCCAACGGAAAGCCCTTTATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCAGTTC 837
Db 12115 GGATCCAAAGAGCTTTCGGAGTCAATTTTAGATGAAAGAGCAAAGAATGTCCTGAAGCTT 12174
Qy 838 GTCAAGAAGTGTGCTGCCTCCGGAGTGACGGAGAACGGCCCCGAGACGACTGTCAACAAC 897
Db 12175 GTGAAACATTCCTGGCAGAATACAGAAGCAGAAAATCATTGTGAGCTAAATAATGATTCC 12234
Qy 898 ACCCCCGAAACGGCAGCTCTCCTCCGGAAGGTTGGCAACGAGGGCATCGTGCTGCTGAAG 957
Db 12235 TCTTGTTTAAGAGAAGC---ACTTAAAAAATTTGCAAGCCAGTCCATTGTTCTCTTGAAA 12291
Qy 958 AACGAGAACAACGTTCTGCCCTTGAGCAAGAAGAAGAAGACGCTGATTGTCGGCCCCAAC 1017
Db 12292 AATAAAAAGAAACTTTTGCCATTGTCAAGGAAGGGCACGTTTGCAGTTATTGGACCTAAC 12351
Qy 1018 GCCAAGCAGGCCACATACCACGGCGGAGGCTCTGCCGCACTCAGGGCCTACTACGCAGTC 1077
Db 12352 GCAAAGGTTTGTAATTACAGCGGAGGTGGATCGGCAAATTTGAAGCCTTATTATACTGTG 12411
Qy 1078 ACTCCCTTTGACGGCCTCAGCAAGCAGCTCGAGACGCCGCCATCGTACACCGTCGGCGCC 1137
Db 12412 AGTATGTATGATGGCATAGCAGCGAAAATAGACGGTGTGCCAGAGTATGCTCTTGGTTGC 12471
Qy 1138 TACACC------ACCGTTCCTCCCATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCT 1191
Db 12472 CATAACTATCTAAACCTTCCAAATATTGCAAATTTATTAGTCAACCCTAGAACCGGCAAG 12531
Qy 1192 CCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCCCTGGTACCCCTAACCGCCAGCACATT 1251
Db 12532 CATGGATACGTTGCGAAGTTTTATCTTGAACCTGCAACCTCCGAAAATAGAACTTTGATA 12591
Qy 1252 GACGAGCTCTTCTTCACCAAGACGGACATGCACCTGGTGGACTACTACCACCCCAA---G 1308
Db 12592 GACGATTACGATTTAGAAGATGGAGTTGTTCGTTTTTACGATTATTGTAACGACAAAATG 12651
Qy 1309 GCGGCAGACACGTGGTACGCCGACATGGAGGGCACGTACACCGCCGACGAGGACTGCACC 1368
Db 12652 AAGGATGGCTATTTCTATATAGACATTGAAGGTTATTTAATTCCTGACGAGGACGCTGTT 12711
Qy 1369 TACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGGCGTACGTAGACGACCAGCTCGTC 1428
Db 12712 TATGAATTTGGGATATCTGTGTTTGGAACGGCTCTGCTGTTTATTGATGATGTCTTATTA 12771
Qy 1429 GTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAG 1488
Db 12772 ATCGACAACAAGACAAAGCAAACCCCTACGAATCATACTTTTGAATTTGGAACAATAGAA 12831
Qy 1489 GAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGTACAAGTTCAAGATCGAGTTCGGC 1548
Db 12832 GAGAGAAATTCTATTTATTTAAAGAAGGGGAGAAAGTATAATGTACGGGTCGAGTATGGA 12891
Qy 1549 TCCGCACCCACCTACACCCTCAAGGGCGACACCATCGTCCCCGGCCACGGCTCCCTCCGC 1608
Db 12892 AGTGCAGCCACTTACACGCTGTCAACTAATCTGTCTCCCAGTACGGGAGGAAGATACTCT 12951
Qy 1609 GTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCC 1668
Db 12952 ATTGGCTGTGTTAAAGTTATTGATCCTGAAACGGAGATAGATTACGCAGTGAGGGTTGCC 13011
Qy 1669 AAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGC 1728
Db 13012 AAATCAGTTGATTGTGTGATTTTATGTGTGGGTTTGACAGCAGAATGGGAAACGAAGGT 13071
Qy 1729 GCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCC 1788
Db 13072 GAAGATAGAAAAACAATGACTTTACCATCGTTATCTGACAAACTGGTTTATTCTATTCTG 13131
Qy 1789 GCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATGCCCTGG 1848
Db 13132 CAATCCAATCCCAATACAGTAGTTGTGACTCAATCAGGCACTCCAATTGAGATGCCATGG 13191
Qy 1849 CTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCC 1908
Db 13192 ATCTCTGAAGCCCACACCTTATTGCACATTTGGTACAACGGTAACGAATTAGGAAATGCT 13251
Qy 1909 ATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCC 1968
Db 13252 TTAGCAAACATTATTTTTGGAGAGCAAAACCCTTGTGGAAAATTGCCAATAACATTCCCT 13311
Qy 1969 AAGCGCCTGCAGGACAACCCCGCGTTTCTCAACTTCCGCACCGAGGCCGCGCGCACGCTG 2028
Db 13312 AAAAAACTGAAGGATAATCCTGCCTATCTTTCATTTCGTTCAAGCCGTGGCCATTGTGTT 13371
Qy 2029 TACGGCGAGGACGTCTACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAAT 2088
Db 13372 TATGGAGAAGACGTCTTTGTTGGGTATAAGTATTATGAAGCCGTAGAAAGGGAGGTCTTG 13431
Qy 2089 TTCCCCTTTGGCCACGGCCTGTCCTACACCACTTTTGCCTTTTCCAATCTCTCCGTGTCT 2148
Db 13432 TTTCCATTTGGATATGGGCTTTCATATACTACTTTTGAATTATCCAATCTTTATTTGAAG 13491
Qy 2149 CACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGC 2208
Db 13492 AATTGTGGTGAAAGGTTAAGGATCGATTTGGAGATTTCGAATACTGGCCCCATGTCTGGT 13551
Qy 2209 GCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTC 2268
Db 13552 GCAGAAATTATACAAGTATATATTTCTCAAATC---GTTCGTTCAGTTAATCGACCTGTT 13608
Qy 2269 AAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCAGCCCGGCGAGACGAAGGCGGTGACA 2328
Db 13609 AAAGAACTAAAAGAGTTTTCTAAAGTGGTTTTATGTCCTAAAGAGACCAAATTAATTAGA 13668
Qy 2329 ATCGAGGAGCAGGAGAAGTACGTCGCTGCGTATTTTGATGAGGAGCGGGATCAGTGGTGT 2388
Db 13669 ATTGAGTTGGATATAAAGTATGCTACTTCTTTTTATGATGAACTTAATGAAAAATGGTGT 13728
Qy 2389 GTCGAAAAGGGTGACTATGAGGTTATCGTGAGCGACAGCAGCGCAG 2434
Db 13729 TCTGAAGAGGGCGAGTACAACGTTCTTGTCGGAACCAGCAGTAAAG 13774

```
RESULT 5
SPU33010
LOCUS       SPU33010                85837 bp    DNA     linear   PLN 14-NOV-1995
DEFINITION  Schizosaccharomyces pombe cosmids 359, 1198 and 1683.
ACCESSION   U33010
VERSION     U33010.1  GI:1063617
KEYWORDS    .
SOURCE      Schizosaccharomyces pombe.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 85837)
  AUTHORS   McCombie,W.R.
  TITLE     Sequence analysis of a region of the fission yeast genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 85837)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-1995) W. Richard McCombie, Cold Spring Harbor
            Lab, 1 Bungtown Rd, P.O. Box 100, Cold Spring Harbor, NY 11754, USA
FEATURES             Location/Qualifiers
     source          1..85837
                     /organism="Schizosaccharomyces pombe"
                     /db_xref="taxon:4896"
                     /chromosome="2"
                     /clone="cosmids 359 (partial), 1198 and 1683"
                     /clone_lib="Mizukami, et al., 1993. Cell 73:121-132."
BASE COUNT    27560 a  14781 c  15395 g  28101 t
ORIGIN

  Query Match             15.7%;  Score 393.6;  DB 8;  Length 85837;
  Best Local Similarity   49.0%;  Pred. No. 7.2e-39;
  Matches 1199;  Conservative    0;  Mismatches 1229;  Indels   18;  Gaps    5;

Qy          1 ATGGCTGATATTGATGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGAT 60
Db      29116 ATGATGGAACATGATGTTGAAGATTTGATCAATCAATTGGATATAAGTGAGAAGGCTATG 29175

Qy         61 CTGCTGGCTGGTATCGACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCT 120
Db      29176 CTGCTTTCCGGCACTGACCTTTGGCATACTGCAGCAATTCCTCGTCTTAACATTCCTTCC 29235

Qy        121 CTCCGCTTTACAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCTTCAATGGCGTCCCT 180
Db      29236 ATTAGATTATCAGACGGCCCTAATGGTATCCGTGGAACTAGCTTCTTCAATTCTTCACCA 29295

Qy        181 GCGGCCTGCTTCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAA 240
Db      29296 TCCGCTTGTTTCCCTGTGGGACAGCGCTAGGGGCTACTTTCGACAAAAAGTTACTATTC 29355

Qy        241 GAGGCAGGTAAGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGC 300
Db      29356 GAAGTTGGTGAATATTTAGCAGAAGAAGCAAAAGCGAAAGGTGTTAGTGTGGTTTTGGGT 29415

Qy        301 CCGACTATCAACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAG 360
Db      29416 CCAACGGTAAATATCCATCGTGGACCCTTAAATGGTAGGGGTTTTGAATCATTTTCTGAA 29475

Qy        361 GATCCGTTCCTGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGA 420
Db      29476 GATTCTACTTTATCAGGTCTTGCAGCTAGTTATGTTATTCTTGGATTGCAAAGTAAAAAC 29535

Qy        421 GTGCAGGCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATGATGGTG 480
Db      29536 GTCCAAGCATGCATTAAGCATTTTGTGTGTAATGATATGGAAGATGAAAGGAACTCTGTC 29595

Qy        481 CAGAGCATCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATTGCT 540
Db      29596 AGTATTGACGTCTCACAAAGGGCACTTAGAGAAGTATATCTTATGCCATTTCAGTTGGCA 29655

Qy        541 GTGCGAGACTCCCAGCCGGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCGTGTCG 600
Db      29656 TGCAAATATTCAAATTTCAAATCACTTATGACTTCTTACAACAAGGTGAATGGTGAACAC 29715

Qy        601 TGCAGCGAGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGC 660
Db      29716 GTTTCCCAATCTCGAATTTTATTAGACAATATTTTAAGAAAAGAGTGGGAGTGGAAAGGT 29775

Qy        661 CTAATCATGAGCGACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAGGCCTC 720
Db      29776 ACTATAATATCTGATTGGTTTGGAACTTATTCGTTGAAGAAAGCTATAGATGCAGGTTTG 29835

Qy        721 GACCTCGAGATGCCCGGACCTCCACGCTTCCGAGGAG---AAACACTCAAGTTCAACGTC 777
Db      29836 GATTTGGAGATGCCAGGAAAGCCGAGATTTCGAAATGTCAACACTATTCAACACTTAGTA 29895

Qy        778 TCCAACGGAAAGCCCTTTATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCAGTTC 837
Db      29896 GGATCCAAAGAGCTTTCGGAGTCAATTTTAGATGAAAGAGCAAAGAATGTCCTGAAGCTT 29955

Qy        838 GTCAAGAAGTGTGCTGCCTCCGGAGTGACGGAGAACGGCCCCGAGACGACTGTCAACAAC 897
Db      29956 GTGAAACATTCCTGGCAGAATACAGAAGCAGAAAATCATTGTGAGCTAAATAATGATTCC 30015

Qy        898 ACCCCGAAACGGCAGCTCTCCTCCGGAAGGTTGGCAACGAGGGCATCGTGCTGCTGAAG 957
Db      30016 TCTTGTTTAAGAGAAGC---ACTTAAAAAATTTGCAAGCCAGTCCATTGTTCTCTTGAAA 30072

Qy        958 AACGAGAACAACGTTCTGCCCTTGAGCAAGAAGAAGAAGACGCTGATTGTCGGCCCCAAC 1017
Db      30073 AATAAAAAGAAACTTTTGCCATTGTCAAGGAAGGGCACGTTTGCAGTTATTGGACCTAAC 30132

Qy       1018 GCCAAGCAGGCCACATACCACGGCGGAGGCTCTGCCGCACTCAGGGCCTACTACGCAGTC 1077
Db      30133 GCAAAGGTTTGTAATTACAGCGGAGGTGGATCGGCAAATTTGAAGCCTTATTATACTGTG 30192

Qy       1078 ACTCCCTTTGACGGCCTCAGCAAGCAGCTCGAGACGCCGCCATCGTACACCGTCGGCGCC 1137
Db      30193 AGTATGTATGATGGCATAGCAGCGAAAATAGACGGTGTGCCAGAGTATGCTCTTGGTTGC 30252

Qy       1138 TACACC------ACCGTTCCTCCCATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCT 1191
Db      30253 CATAACTATCTAAACCTTCCAAATATTGCAAATTTATTAGTCAACCCTAGAACCGGCAAG 30312

Qy       1192 CCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCCCTGGTACCCCTAACCGCCAGCACATT 1251
Db      30313 CATGGATACGTTGCGAAGTTTTATCTTGAACCTGCAACCTCCGAAAATAGAACTTTGATA 30372

Qy       1252 GACGAGCTCTTCTTCACCAAGACGGACATGCACCTGGTGGACTACTACCACCCCAA---G 1308
Db      30373 GACGATTACGATTTAGAAGATGGAGTTGTTCGTTTTTACGATTATTGTAACGACAAAATG 30432

Qy       1309 GCGGCAGACACGTGGTACGCCGACATGGAGGGCACGTACACCGCCGACGAGGACTGCACC 1368
Db      30433 AAGGATGGCTATTTCTATATAGACATTGAAGGTTATTTAATTCCTGACGAGGACGCTGTT 30492

Qy       1369 TACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGGCGTACGTAGACGACCAGCTCGTC 1428
Db      30493 TATGAATTTGGGATATCTGTGTTTGGAACGGCTCTGCTGTTTATTGATGATGTCTTATTA 30552

Qy       1429 GTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAG 1488
Db      30553 ATCGACAACAAGACAAAGCAAACCCCTACGAATCATACTTTTGAATTTGGAACAATAGAA 30612

Qy       1489 GAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGTACAAGTTCAAGATCGAGTTCGGC 1548
Db      30613 GAGAGAAATTCTATTTATTTAAAGAAGGGGAGAAAGTATAATGTACGGGTCGAGTATGGA 30672

Qy       1549 TCCGCACCCACCTACACCCTCAAGGGCGACACCATCGTCCCCGGCCACGGCTCCCTCCGC 1608
Db      30673 AGTGCAGCCACTTACACGCTGTCAACTAATCTGTCTCCCAGTACGGGAGGAAGATACTCT 30732

Qy       1609 GTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCC 1668
Db      30733 ATTGGCTGTGTTAAAGTTATTGATCCTGAAACGGAGATAGATTACGCAGTGAGGGTTGCC 30792
```

```
Qy   1669 AAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGC 1728
          ||       ||    || ||  | || | |||  | |  || || ||||| || || ||
Db  30793 AAATCAGTTGATTGTGTGATTTTATGTGTGGGTTTGACAGCAGAATGGGAAACGGAAGGT 30852

Qy   1729 GCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCC 1788
          |  ||  |    |  ||||   | ||      |    ||| | ||  ||        |
Db  30853 GAAGATAGAAAAACAATGACTTTACCATCGTTATCTGACAAACTGGTTTATTCTATTCTG 30912

Qy   1789 GCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATGCCCTGG 1848
              | || || || || || || || | ||  | ||||| ||    |||||||| |||
Db  30913 CAATCCAATCCCAATACAGTAGTTGTGACTCAATCAGGCACTCCAATTGAGATGCCATGG 30972

Qy   1849 CTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCC 1908
           ||   |  | | | || |  | ||    ||||||  ||| |||||    || ||  |
Db  30973 ATCTCTGAAGCCCACACCTTATTGCACATTTGGTACAACGGTAACGAATTAGGAAATGCT 31032

Qy   1909 ATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCC 1968
          | ||  || |  | ||||| ||  | ||||| |  || ||  || |  | |  |||||
Db  31033 TTAGCAAACATTATTTTTGGAGAGCAAAACCCTTGTGGAAAATTGCCAATAACATTCCCT 31092

Qy   1969 AAGCGCCTGCAGGACAACCCCGCGTTTCTCAACTTCCGCACCGAGGCCGGGCGCACGCTG 2028
          ||    ||| |||| || || || | |||    || ||  |     || |      |
Db  31093 AAAAAACTGAAGGATAATCCTGCCTATCTTTCATTTCGTTCAAGCCGTGGCCATTGTGTT 31152

Qy   2029 TACGGCGAGGACGTCTACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAAT 2088
          || || || |||||||  || ||||| | ||| || ||    |  || | ||| |||
Db  31153 TATGGAGAAGACGTCTTTGTTGGGTATAAGTATTATGAAGCCGTAGAAAGGGAGGTCTTG 31212

Qy   2089 TTCCCCTTTGGCCACGGCCTGTCCTACACCACTTTTGCCTTTTCCAATCTCTCCGTGTCT 2148
          || || |||||  | || || || || || |||||||  || ||||||| |   ||
Db  31213 TTTCCATTTGGATATGGGCTTTCATATACTACTTTTGAATTATCCAATCTTTATTTGAAG 31272

Qy   2149 CACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGC 2208
           |    |  |  | | | ||  |      |      |   ||| || ||| || || | ||
Db  31273 AATTGTGGTGAAAGGTTAAGGATCGATTTGGAGATTTCGAATACTGGCCCCATGTCTGGT 31332

Qy   2209 GCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTC 2268
          ||| |  |    ||  | ||  |    |   ||   |          |||| || || ||
Db  31333 GCAGAAATTATACAAGTATATATTTCTCAAATC---GTTCGTTCAGTTAATCGACCTGTT 31389

Qy   2269 AAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCAGCCCGGCGAGACGAAGGCGGTGACA 2328
          || || || || |  ||  | || || |    |    ||    ||||| ||      | | |
Db  31390 AAAGAACTAAAAGAGTTTTCTAAAGTGGTTTTATGTCCTAAAGAGACCAAATTAATTAGA 31449

Qy   2329 ATCGAGGAGCAGGAGAAGTACGTCGCTGCGTATTTTGATGAGGAGCGGGATCAGTGGTGT 2388
          || |||  | |    |||||  |   || | | || ||||||        ||  | ||||||
Db  31450 ATTGAGTTGGATATAAAGTATGCTACTTCTTTTTATGATGAACTTAATGAAAAATGGTGT 31509

Qy   2389 GTCGAAAAGGGTGACTATGAGGTTATCGTGAGCGACAGCAGCGCAG 2434
            ||| |||| || ||  | ||| | ||  |   ||||||   ||
Db  31510 TCTGAAGAGGGCGAGTACAACGTTCTTGTCGGAACCAGCAGTAAAG 31555
```

RESULT 6
AF329731

```
LOCUS       AF329731               2793 bp    mRNA    linear   PLN 28-JAN-2001
DEFINITION  Volvariella volvacea beta-glucosidase mRNA, complete cds.
ACCESSION   AF329731
VERSION     AF329731.1  GI:12584216
KEYWORDS    .
SOURCE      Volvariella volvacea.
  ORGANISM  Volvariella volvacea
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Pluteaceae; Volvariella.
REFERENCE   1  (bases 1 to 2793)
  AUTHORS   Ding,S., Ge,W. and Buswell,J.A.
  TITLE     Molecular cloning, characterization and expression of a
            beta-glucosidase gene from the edible straw mushroom, Volvariella
            volvacea
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2793)
  AUTHORS   Ding,S., Ge,W. and Buswell,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-2000) Biology, The Chinese University of Hong
            Kong, Shatin, N.T., Hong Kong, China
FEATURES             Location/Qualifiers
     source          1. .2793
                     /organism="Volvariella volvacea"
                     /strain="V14"
                     /db_xref="taxon:36659"
     CDS             41. .2629
                     /EC_number="3.2.1.21"
                     /codon_start=1
                     /product="beta-glucosidase"
                     /protein_id="AAG59831.1"
                     /db_xref="GI:12584217"
                     /translation="MPPSDFAKANIDEIVEQLTLDEAISLTAGVGFWHTHAIERLGVP
                     AVKVSDGPNGIRGNHFFMGTPAKCLPSSTALGATWDPEVVEEVGLKLLAPEAKLRAAS
                     LVLAPTSNIQRNPLGGRSFESFSEDPYLSGIISASYVNGVQKGGIGATIKHFVGNDKE
                     DDRQGYDSIISERALREIYLLPFMLTQKYAAPWAIMTAYNRVNGVHVAEDPFLLKQVL
                     RNEWKYKGLIMSDWFGMYSVDHGIKAGLDLEMPGINKWRTLDLVNRTIQARKLTPRDI
                     KDRARVVLELVKKCAQGAPEILDGDGEERTVELESDKLLMRRIASESIVLLKNDNVLP
                     LKPEGGAIKKIAVVGGNAKAQVLSGGGSAALKASYFISPYDGIKAALEPHGVEVTFSE
                     GARAYKTLPTLEWDLETETGERGWIGTWHTHESDDSMTALDQPFIAPRLVDETRIFIS
                     TSYPKGITKRWTMRLKGYLKPREKDTNFEFGLIAAGRAKLWVDGQLVIDNWTRQRRGE
                     AFFGSGSQEETGVYLLKAGKKHEIYVEYCNVRAPADGDEDEAIMDSNPGVRLGGAEVA
                     NADDLLSEAVKLASEADAVIAVVGLNADWETEGNDRRTLALPGRTDELVEKVAKVNSK
                     TVVVTQAGSAITLPWLDSVAAVVHAWYLGNATGDAIADVLFGKQNPSGKLSLTFPKRL
                     EDVPSHGHFGSENGKVRYAEDLFVGYKHYHHRNIEPLFPFGFGLSYTTFSLSNLQLSA
                     PVIDHATSSFSLTATLSITNTGPVTGSEVAQLYVSYPETSELTHAPLQLRAFKKVKDL
                     QPGETREVKLELDKYAVSYWNDRYQTWAVENGEYEIKVGNSSVSKDLTLVQRFVVKDG
                     FEWRGI"
BASE COUNT      697 a    696 c    737 g    663 t
ORIGIN
```

Query Match 15.7%; Score 392.2; DB 8; Length 2793;
Best Local Similarity 50.3%; Pred. No. 2.6e-38;
Matches 1258; Conservative 0; Mismatches 1168; Indels 75; Gaps 9;

```
Qy     17 TTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGATCTGCTGGCTGGTATCG 76
          |||| |  ||| |  || |  | || || |  |||    ||  ||||   |||||| |||
Db     72 TTGATGAAATCGTCGAGCAATTGACGCTCGATGAGGCCATCAGTCTGACCGCTGGTGTCG 131

Qy     77 ACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCTCTCCGCTTTACAGATG 136
          ||||||||  ||  | ||  |        |  || || ||  || |    | |  ||||
Db    132 GTTTCTGGCATACGCATGCGATTGAGCGTCTCGGTGTGCCAGCTGTAAAGGTCAGCGATG 191

Qy    137 GCCCCAACGGCGTAAGAGGGACCAAGTTCTTCAATGGCGTCCCTGCGGCCTGCTTCCCTT 196
          | || |||||| |  |||| | | | |||||||  ||   ||||||    ||  | ||||
Db    192 GACCAAACGGCATTCGAGGCAACCACTTCTTCATGGGGACCCCTGCAAAATGTCTACCTT 251

Qy    197 GCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGG---TAAGA 253
          |   ||  | || ||  | || |   |||      |  |||| |||| ||    |||
Db    252 CCTCAACCGCTCTTGGCGCAACTTGGGACCCTGAAGTCGTCGAGGAGGTCAGGTTGAAGC 311

Qy    254 TGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACA 313
          |  | | |   || || |           ||   | |  | || |  || ||| |||||
Db    312 TCCTAGCCCCTGAAGCTAAGCTGCGAGCGGCTTCTCTAGTTCTTGCTCCTACTTCCAACA 371

Qy    314 TGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGG 373
          | ||| |   ||| || || || ||| | ||||||     |   |||||||| | |||
Db    372 TCCAAAGAAACCCCCTGGGAGGCCGTAGTTTCGAGAGTTTCTCTGAGGATCCCTACCTTT 431

Qy    374 CGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGA 433
          ||||  |     | ||| |   | ||   ||  | ||||    |||| |    ||||| |
Db    432 CGGGTATCATTTCGGCGTCCTACGTCAATGGTGTGCAGAAGGGTGGAATTGGTGCTACTA 491

Qy    434 TCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCA 493
          |||| || ||| |  | || ||  ||||||||  ||           | || ||  |||
Db    492 TCAAACATTTTGTCGGAAACGACAAGGAGGACGATCGTCAAGGATACGATAGTATTATCA 551

Qy    494 CGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCC 553
            || || |||||||| ||||||||   |||||| ||   | |  ||        | |
Db    552 GCGAACGTGCTCTCCGCGAAATCTATCTACTCCCTTTTATGTTAACTCAAAAGTATGCTG 611
```

```
Qy    554 AGCCGGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACC 613
Db    612 CCCCATGGGCAATCATGACCGCCTACAACCGAGTGAACGGCGTTCATGTAGCGGAAGACC 671

Qy    614 CTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCG 673
Db    672 CCTTCCTGCTCAAGCAGGTCCTACGTAATGAGTGGAAATACAAGGGATTGATCATGAGTG 731

Qy    674 ACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGC 733
Db    732 ACTGGTTCGGCATGTACAGCGTTGATCATGGTATCAAGGCCGGTCTAGACCTCGAAATGC 791

Qy    734 CCGGACCTCCACGCTTCCGAGGA---GAAACACTCAAGTTCAACGTCTCCAACGGAAAGC 790
Db    792 CTGGCATCAATAAATGGCGTACACTTGATCTAGTCAACCGTACTATTCAAGCTCGCAAAC 851

Qy    791 CCTTTATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTG 850
Db    852 TCACCCCACGAGACATCAAGGACCGAGCACGAGTTGTATTAGAGTTAGTCAAGAAGTGTG 911

Qy    851 CT------GCCTCCGGAGTGACGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCG 904
Db    912 CTCAAGGAGCCCCAGAGATTCTCGATGGCGATGGCGAGGAGCGTACAGTCGAGCTTGAAT 971

Qy    905 AAACGGCAGCTCTCCTCCGGAAGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGA 964
Db    972 CCGACAAGCTGCTCATGCGGCGCATCGCATCTGAGTCCATCGTCCTACTCAAAAATGATA 1031

Qy    965 ACAACGTTCTGCCCT---------TGAGCAAGAAGAAGAAGACGCTGATTGTCGGCCCCA 1015
Db   1032 ACGTTCTCCCTCTTAAACCTGAAGGAGGCGCGATCAAGAAGATCGCCGTCGTTGGTGGCA 1091

Qy   1016 ACGCCAAGCAGGCCACATACCACGGCGGAGGCTCTGCCGCACTCAGGGCCTACTACGCAG 1075
Db   1092 ATGCAAAGGCGCAAGTCCTGTCCGGCGGAGGGTCTGCGGCGCTGAAGGCGTCATACTTCA 1151

Qy   1076 TCACTCCCTTTGACGGCCTCAGCAAGCAGCTCGAGACGCCGCCATCGTACACC------- 1128
Db   1152 TCTCACCCTACGACGGTATTAAGGCCGCCCTCGAACCTCATGGCGTAGAGGTCACATTCA 1211

Qy   1129 --------GTCGGCGCCTACACCACCGTTCCTCCCATTCTAGGCGAGCAGTGCCTCACGC 1180
Db   1212 GTGAGGGTGCTCGAGCCTATAAAACACTACCCACCCTCGAATGGGATCTGGAAACCGAGA 1271

Qy   1181 CCGACGGCGCTCCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCCCTGGTA--------- 1231
Db   1272 CCGGCGAACGTGGATGGATTGGAACGTGGCACACCCACGAGTCCGACGACAGCATGACGG 1331

Qy   1232 CCCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGACGGACATGCACCTGGTGG 1291
Db   1332 CTCTTGACCAGCCATTCATCGCTCCCCGATTGGTTGACGAAACCAGAATCTTCATCAGCA 1391

Qy   1292 ACTACTACCACCCCAAGGCGGCAGACACGTGGTACGCCGACATGGAGGG---CACGTACA 1348
Db   1392 CGTCGTATCCCAAGGGAATTACAAAGAGGTGGACGATGAGACTGAAGGGCTACCTGAAGC 1451

Qy   1349 CCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGGCGT 1408
Db   1452 CAAGAGAGAAGGACACCAACTTCGAATTTGGTCTTATTGCTGCTGGTCGGGCTAAGCTCT 1511

Qy   1409 ACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCTTCT 1468
Db   1512 GGGTTGACGGACAACTGGTAATCGACAACTGGACCCGTCAACGTCGTGGTGAAGCATTCT 1571

Qy   1469 TCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGTACA 1528
Db   1572 TTGGCAGTGGTAGTCAGGAGGAAACTGGGGTCTATCTCCTCAAGGCTGGCAAGAAGCATG 1631

Qy   1529 AGTTCAAGATCGAGTTCGGCTC------CGCACCCACCTACACCCTCAAGGGCGACACCA 1582
Db   1632 AAATCTATGTCGAGTACTGTAATGTGCGCGCTCCGGCTGATGGTGATGAAGATGAGGCCA 1691

Qy   1583 TCGTCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCG 1642

Db   1692 TTATGGATAGTAACCCTGGTGTTCGCCTTGGTGGTGCAGAGGTTGCCAACGCAGATGATC 1751

Qy   1643 AAATCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCC 1702
Db   1752 TCCTATCTGAGGCTGTCAAACTTGCTTCAGAGGCTGATGCCGTCATTGCGGTCGTTGGCC 1811

Qy   1703 TTAACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGC 1762
Db   1812 TCAACGCAGATTGGGAGACTGAGGGTAATGACAGAAGGACATTGGCTCTTCCTGGGCGAA 1871

Qy   1763 TGGACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCCTCATGCAGA 1822
Db   1872 CGGATGAATTAGTTGAGAAGGTCGCCAAGGTCAACTCGAAGACTGTTGTAGTTACCCAAG 1931

Qy   1823 CGGGCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGT 1882
Db   1932 CTGGTTCAGCTATAACTTTGCCATGGCTCGACTCTGTCGCTGCAGTCGTTCACGCATGGT 1991

Qy   1883 ACGGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCT 1942
Db   1992 ATCTTGGTAATGCTACTGGAGATGCTATTGCGGACGTGCTGTTCGGCAAGCAGAATCCAT 2051

Qy   1943 CGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAACCCCGCGTTTCTCAACT 2002
Db   2052 CTGGTAAATTGTCGTTGACATTCCCTAAGAGGTTGGAGGACGTGCCTTCGCATGGACACT 2111

Qy   2003 TCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTCGGGTACAGGTACT 2062
Db   2112 TCGGTTCCGAGAACGGCAAGGTACGATATGCCGAAGACTTGTTCGTTGGTTACAAGCACT 2171

Qy   2063 ACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACACCACTT 2122
Db   2172 ACCACCACCGCAACATCGAGCCTCTTTTCCCCTTCGGCTTTGGTCTCTCCTACACCACAT 2231

Qy   2123 TTGCCTTTTCCAATCTCTCCGTGT----------CTCACAAGGACGGCA 2161
Db   2232 TCTCCCTCTCCAATCTTCAACTCTCGGCACCAGTCATTGATCATGCAACCTCATCCTTCT 2291

Qy   2162 AGCTGAGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCC 2221
Db   2292 CTCTCACCGCAACCCTTTCCATCACCAACACCGGTCCCGTTACCGGCTCCGAAGTCGCCC 2351

Qy   2222 AGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGG 2281
Db   2352 AACTCTACGTCAGCTATCCCGAAACCTCGGAACTGACACACGCACCACTACAGCTGAGGG 2411

Qy   2282 GCTTCGCAAAGGTCGAACTGCAGCCCGGCGAGACGAAGGCGGTGACAATCGAGGAGCAGG 2341
Db   2412 CGTTCAAGAAGGTGAAGGACTTGCAGCCTGGCGAGACTCGCGAGGTGAAGCTGGAGCTGG 2471

Qy   2342 AGAAGTACGTCGCTGCGTATTTTGATGAGGAGCGGGATCAGTGGTGTGTCGAAAAGGGTG 2401
Db   2472 ACAAATATGCTGTGTCGTATTGGAATGACAGGTATCAGACCTGGGCTGTGGAGAACGGGG 2531

Qy   2402 ACTATGAGGTTATCGTGAGCGACAGCAGCGCAGCGAAGGAT 2442
Db   2532 AGTATGAGATCAAGGTTGGAAATAGCAGCGTGTCAAAGGAT 2572
```

RESULT 7
AF005277

LOCUS	AF005277	3578 bp DNA linear BCT 02-MAR-1998
DEFINITION	Cellulomonas biazotea cellobiase (cba) gene, complete cds.	
ACCESSION	AF005277	
VERSION	AF005277.1 GI:2921739	
KEYWORDS	.	
SOURCE	Cellulomonas biazotea.	
ORGANISM	Cellulomonas biazotea Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococcineae; Cellulomonadaceae; Cellulomonas.	
REFERENCE	1 (bases 1 to 3578)	
AUTHORS	Wong,W.K., Ali,A., Chan,W.K., Ho,V. and Lee,N.T.	

```
  TITLE     The cloning, expression and characterization of a cellobiase gene
            encoding a secretory enzyme from Cellulomonas biazotea
  JOURNAL   Gene 207 (1), 79-86 (1998)
  MEDLINE   98172737
   PUBMED   9511746
REFERENCE   2  (bases 1 to 3578)
  AUTHORS   Wong,W.-K.R., Ali,A., Chan,R.W.K., Ho,V. and Lee,N.T.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAY-1997) Biochemistry, The Hong Kong University of
            Science and Technology, Clear Water Bay Road, Kowloon, Hong Kong
FEATURES             Location/Qualifiers
     source          1. .3578
                     /organism="Cellulomonas biazotea"
                     /strain="ATCC 486"
                     /db_xref="ATCC:486"
                     /db_xref="taxon:1709"
     gene            327. .2813
                     /gene="cba"
     CDS             327. .2813
                     /gene="cba"
                     /function="hydrolyzing beta-D-glucoside and cellobiose"
                     /note="one type of cellulase"
                     /codon_start=1
                     /transl_table=11
                     /product="cellobiase"
                     /protein_id="AAC38196.1"
                     /db_xref="GI:2921740"
                     /translation="MTSQTALDPAALVASLPLETKVRLLTGATAFTLAPEESIGLGEV
                     RLSDGPTGVRGLKFSGGRTVALFPNATLLASAWSEESTTEVGRLLAEEALAQQIHVVL
                     GPTINLHRSVLGGRLFEAYSEDPLLTGRLAAAYVRGLQDLGVGACLKHLVANESETER
                     NTMNSVVDPATLRELYLLPFEIAVDESDPWSVMAAYNDVNGVPATEHHHVVNEVLKGE
                     WGYTGLVMSDWFATRTAAPAAAGGLDLVMPGPDGPWGDALVAAVRSGELDESVVDDHL
                     RRLLVLAARVGALGDLRDYPDDLPAPDSAVRREQLTRLAAAGMTVLTNADDTLPLARG
                     TRVALVGRHALETIDMGGGSATVNPPYQVSVAEGLTALLGDAVDVVDGVEVRTRPVPA
                     RPGFVVDPDTGRPGLHLTLLAADGTVLDERHDAPSTVMVGFDDDFPQAVARVRFRARV
                     AGEGALEVGAIGVGRWQVTAGGTELAWTLATSGTGFAEEMLAPPTRTDQVHVGSDAVV
                     DATVVLRSSTRSVTVGDADPGTDAGAAAEPLAGVGLFGLVARPAPEAEDDVITRAAAA
                     AAQADVAVVVVGLTEEEETESVDKSTIALPGAQDALVRAVAAAARRTVVVVNAATPVL
                     MPWLDDVDAVLWAGLPGQEGGHAVAAALLGDQEPTGRLVTTFPAADGAAPAWSVTPVD
                     GDLEYTEGRFVGYRGHWADRAPAPAFWLGHGLGYATWEYADATLDTDGDAPAVTVTVT
                     NTGARTSREVVQVYLEPASSDEPVRLVGWADATVDAGASARVTVTADARMWRRWDEAA
                     GGWSRLADGGRLLVARGLGDVRATLALPTA"
BASE COUNT      426 a   1421 c   1306 g    425 t
ORIGIN

  Query Match             14.0%;  Score 350.4;  DB 1;  Length 3578;
  Best Local Similarity   48.7%;  Pred. No. 2.5e-33;
  Matches 1087;  Conservative    0;  Mismatches 1121;  Indels   24;  Gaps    4;

Qy    109 GGAGTCCCCTCTCTCCGCTTTACAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCTTC 168
Db    444 GGCCTCGGCGAGGTCCGCCTGTCCGACGGCCCGACCGGCGTGCGCGGCCTCAAGTTCTCC 503

Qy    169 AATGGCGTCCCTGCGGCCTGCTTCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAA 228
Db    504 GGCGGGCGCACCGTCGCGCTGTTCCCCAACGCGACCCTGCTCGCGTCCGCGTGGAGCGAG 563

Qy    229 ACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCAT 288
Db    564 GAGTCGACGACCGAGGTCGGCCGGCTGCTCGCCGAGGAGGCCCTCGCGCAGCAGATCCAC 623

Qy    289 GTGATCCTCGGCCCGACTATCAACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAG 348
Db    624 GTCGTGCTCGGCCCCACGATCAACCTGCACCGCTCCGTGCTCGGCGGACGCCTGTTCGAG 683

Qy    349 TCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATT 408
Db    684 GCCTACTCCGAGGACCCGCTGCTCACCGGCCGGCTCGCCGCCGCCTACGTGCGCGGCCTG 743

Qy    409 CAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAGG 468
Db    744 CAGGACCTCGGCGTCGGCGCCTGCCTCAAGCACCTCGTCGCCAACGAGTCGGAGACCGAG 803

Qy    469 CGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCG 528
Db    804 CGCAACACCATGAACTCCGTCGTCGACCCCGCGACGCTCCGCGAGCTCTACCTGCTGCCG 863

Qy    529 TTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGTTCATGACGGCGTACAATGGCATC 588
Db    864 TTCGAGATCGCCGTCGACGAGTCCGACCCGTGGTCCGTCATGGCCGCGTACAACGACGTC 923

Qy    589 AATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATGG 648
Db    924 AACGGCGTCCCCGCGACCGAGCACCACCACGTCGTCAACGAGGTCCTCAAGGGCGAGTGG 983

Qy    649 GGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCACATACAGTACCACAGAAGCCGTT 708
Db    984 GGCTACACCGGCCTCGTCATGTCCGACTGGTTCGCGACCCGGACCGCCGCACCCGCCGCC 1043

Qy    709 GTGGCAGGCCTCGACCTCGAGATGCCCGGACCTCCACGCTTCCGAGGAGAAACACTCAAG 768
Db   1044 GCCGGCGGGCTCGACCTCGTCATGCCCGGGCCCGACGGCCCCTGGGGCGACGCGCTCGTC 1103

Qy    769 TTCAACGTCTCCAACGGAAAGCCCTTTATCCACGTCATTGACCAGAGGGCTAGGGAAGTT 828
Db   1104 GCCGCCGTGCGCAGCGGCGAGCTCGACGAGTCCGTCGTCGACGACCACCTGCGCCGCCTG 1163

Qy    829 CTTCAGTTCGTCAAGAAGTGTGCTGCCTCCGGAG---TGACGGAGAACGGCCCCGAGACG 885
Db   1164 CTCGTCCTCGCCGCACGCGTCGGCGCCCTCGGCGACCTGCGGGACTACCCCGACGACCTG 1223

Qy    886 ACTGTCAACAACACCCCCGAAACGGCAGCTCTCCTCCGGAAGGTTGGCAACGAGGGCATC 945
Db   1224 CCCGCCCCCGACAGCGCCGTCCGGCGCGAGCAGCTCACGCGCCTCGCCGCCGCCGGCATG 1283

Qy    946 GTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCAAGAAGAAGAAGACGCTGATT 1005
Db   1284 ACCGTCCTCACCAACGCCGACGACACCCTGCCGCTCGCGCGCGGCACCCGCGTCGCACTC 1343

Qy   1006 GTCGGCCCCAACGCCAAGCAGGCCACATACCACGGCGGAGGCTCTGCCGCACTCAGGGCC 1065
Db   1344 GTCGGGCGGCACGCGCTCGAGACGATCGACATGGGCGGCGGCTCCGCGACCGTCAACCCG 1403

Qy   1066 TACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGCTCGAGACGCCGCCATCGTAC 1125
Db   1404 CCGTACCAGGTGAGCGTCGCCGAGGGCCTCACCGCGCTGCTCGGCGACGCGGTCGACGTC 1463

Qy   1126 ACCGTCGGCGCCTACACCACCGTTCCTCCCATTCTAGGCGAGCAGTGCCTCACGCCCGAC 1185
Db   1464 GTCGACGGCGTCGAGGTCCGCACCCGCCCCGTGCCCGCGCGGCCCGGGTTCGTCGTGGAC 1523

Qy   1186 GGCGCTCCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCCCTGGTACCCCTAACCGCCAG 1245
Db   1524 CCCGACACCGGCCGGCCCGGCCTGCACCTCACGCTCCTCGCCGCCGACGGCACCGTCCTC 1583

Qy   1246 CACATTGACGAGCTCTTCTTCACCAAGACGGACATGCACCTGGTGGACTACTACCACCCC 1305
Db   1584 GACGAGCGGCACGACGCGCCGTCGACCGTCATGGTCGGCTTCGACGACGACTTCCCGCAG 1643

Qy   1306 AAGGCGGCAGACACGTGGTACGCCGACATGGAGGGCACGTACACCGCCGACGAGGACTGC 1365
Db   1644 GCCGTCGCCCGCGTCCGGTTCCGTGCCCGCGTCGCCGGCGAAGGCGCGCTCGAGGTCGGC 1703

Qy   1366 ACCTACGAGCTCGGCC---------TCGTCGTCTGCGGCACGGCAAAGGCGTACGTAGAC 1416
Db   1704 GCGATCGGCGTCGGACGGTGGCAGGTGACCGCAGGCGGCACCGAGCTCGCCTGGACGCTC 1763

Qy   1417 GACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCTTCTTCGGCTCC 1476
Db   1764 GCGACGTCCGGCACCGGGTTCGCCGAGGAGATGCTCGCGCCGCCGACCCGCACCGACCAG 1823

Qy   1477 GCCACCCGCGAG---GAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGTACAAGTTC 1533
Db   1824 GTGCACGTGGGGTCCGACGCGGTCGTCGACGCGACCGTCGTGCTGCGCTCGTCCACCCGC 1883

Qy   1534 AAGATCGAGTTCGGCTCCGCACCCACCTACACCCTCAAGGGCGACACCATCGTCCCCGGC 1593
```

```
Db   1884 AGCGTGACCGTCGGCGACGCCGACCCCGGCACCGACGCGGGCGCCGCCGCCGAGCCGCTC 1943

Qy   1594 CACGGC---------TCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAA 1644

Db   1944 GCGGGCGTCGGGCTGTTCGGGCTCGTCGCCCGACCCGCCCCGGAGGCGGAGGACGACGTC 2003

Qy   1645 ATCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTT 1704

Db   2004 ATCACCCGCGCCGCGGCGGCCGCCGCGCAGGCGGACGTCGCCGTCGTCGTCGTGGGGCTG 2063

Qy   1705 AACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTG 1764

Db   2064 ACCGAGGAGGAGGAGACGGAGTCCGTCGACAAGTCCACGATCGCGCTGCCCGGCGCGCAG 2123

Qy   1765 GACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACG 1824

Db   2124 GACGCGCTCGTGCGGGCCGTCGCCGCGGCCGCTCGCCGCACGGTCGTCGTGGTCAACGCC 2183

Qy   1825 GGCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTAC 1884

Db   2184 GCGACGCCGGTGCTCATGCCGTGGCTCGACGACGTCGACGCCGTGCTGTGGGCCGGGCTC 2243

Qy   1885 GGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCG 1944

Db   2244 CCCGGCCAGGAGGGCGGGCACGCCGTGGCCGCCGCGCTGCTCGGCGACCAGGAGCCGACC 2303

Qy   1945 GGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAACCCCGCGTTTCTCAACTTC 2004

Db   2304 GGCCGCCTCGTCACGACGTTCCCCGCCGCCGACGGCGCCGCACCCGCGTGGTCCGTCACG 2363

Qy   2005 CGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTCGGGTACAGGTACTAC 2064

Db   2364 CCCGTCGACGGCGACCTGGAGTACACCGAGGGCCGGTTCGTCGGCTACCGCGGCCACTGG 2423

Qy   2065 GAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACACCACTTTT 2124

Db   2424 GCCGACCGGGCGCCCGCGCCCGCGTTCTGGCTCGGGCACGGCCTCGGGTACGCGACGTGG 2483

Qy   2125 GCCTTTTCCAATCTCTCCGTGTCTCACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTG 2184

Db   2484 GAGTACGCCGACGCGACGCTCGACACGGACGGCGACGCCCCCGCGGTGACCGTCACCGTG 2543

Qy   2185 AAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCAA 2244

Db   2544 ACCAACACCGGCGCGCGCACCAGCCGCGAGGTCGTCCAGGTCTACCTGGAGCCCGCGTCG 2603

Qy   2245 GCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCAG 2304

Db   2604 TCCGACGAGCCCGTGCGCCTCGTCGGCTGGGCCGACGCGACCGTCGACGCCGGGGCGTCC 2663

Qy   2305 CCCGGCGAGACG 2316

Db   2664 GCCCGCGTCACG 2675


RESULT 8
ATUCBG1
LOCUS       ATUCBG1                 3710 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  A.tumefaciens beta-D-glucosidase (cbg-1) gene, complete cds.
ACCESSION   M59852
VERSION     M59852.1  GI:142221
KEYWORDS    beta-D-glucosidase.
SOURCE      A.tumefaciens DNA.
  ORGANISM  Agrobacterium tumefaciens
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Rhizobium.
REFERENCE   1  (bases 1 to 3710)
  AUTHORS   Castle,L.A., Smith,K.D. and Morris,R.O.
  TITLE     Cloning and sequencing of an Agrobacterium tumefaciens
            beta-glucosidase gene involved in modifying a vir-inducing plant
            signal molecule
  JOURNAL   J. Bacteriol. 174 (5), 1478-1486 (1992)
  MEDLINE   92165721
   PUBMED   1537792
FEATURES             Location/Qualifiers
     source          1. .3710
                     /organism="Agrobacterium tumefaciens"
                     /db_xref="taxon:358"
                     /cell_line="B3/73"
                     /tissue_lib="genomic HindIII cosmid"
     gene            169. .2637
                     /gene="cbg1"
     RBS             169. .173
                     /gene="cbg1"
     CDS             181. .2637
                     /gene="cbg1"
                     /EC_number="3.2.1.21"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-D-glucosidase"
                     /protein_id="AAA22082.1"
                     /db_xref="GI:142222"
                     /translation="MIDDILDKMTLEEQVSLLSGADFWTTVAIERLGVPKIKVTDGPN
                     GARGGGSLVGGVKSACFPVAIALGATWDPELIERAGVALGQQAKSKGASVLLAPTVNI
                     HRSGLNGRNFECYSEDPALTAACAVAYINGVQSQGVAATIKHFVANESEIERQTMSSD
                     VDERTLREIYLPPFEEAVKKAGVKAVMSSYNKLNGTYTSENPWLLTKVLREEWGFDGV
                     VMSDWFGSHSTAETINAGLDLEMPGPWRDRGEKLVAAVREGKVKAETVRASARRILLL
                     LERVGAFEKAPDLAEHALDLPEDRALIRQLGAEGAVLLKNDGVLPLAKSSFDQIAVIG
                     PNAASARVMGGGSARIAAHYTVSPLEGIRAALSNANSLRHAVGCNNNRLIDVFSGEMT
                     VEYFKGRGFESRPVHVETVEKGEFFWFDLPSGDLDLADFSARMTATFVPQETGEHIFG
                     MTNAGLARLFVDGELVVDGYDGWTKGENFFGTANSEQRRAVTLGAARRYRVVVEYEAP
                     KASLDGINICALRFGVEKPLGDAGIAEAVETARKSDIVLLLVGREGEWDTEGLDLPDM
                     RLPGRQEELIEAVAETNPNVVVVLQTGGPIEMPWLGKVRAVLQMWYPGQELGNALADV
                     LFGDVEPAGRLPQTFPKALTDNSAITDDPSIYPGQDGHVRYAEGIFVGYRHHDTREIE
                     PLFPFGFGLGYTRFTWGAPQLSGTEMGADGLTVTVDVTNIGDRAGSDVVQLYVHSPNA
                     RVERPFKELRAFAKLKLAPGATGTAVLKIAPRDLAYFDVEAGRFRADAGKYELIVAAS
                     AIDIRASVSIHLPVDHVMEP"
     misc_feature    213. .223
                     /gene="cbg1"
                     /note="active site"
BASE COUNT      718 a   1084 c   1168 g    740 t
ORIGIN

  Query Match             13.7%;  Score 343.4;  DB 1;  Length 3710;
  Best Local Similarity   49.1%;  Pred. No. 1.7e-32;
  Matches 1195;  Conservative    0;  Mismatches 1171;  Indels   69;  Gaps    8;

Qy      17 TTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGATCTGCTGGCTGGTATCG 76

Db     185 TCGACGATATTCTCGATAAGATGACACTGGAGGAGCAGGTGTCGCTGCTCTCGGGCGCGG 244

Qy      77 ACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCTCTCCGCTTTACAGATG 136

Db     245 ATTTCTGGACGACCGTCGCGATCGAGCGGCTCGGCGTGCCGAAGATCAAGGTTACCGACG 304

Qy     137 GCCCCAACGGCGTAAGAGGGACCAAGT---TCTTCAATGGCGTCCCTGCGGCCTGCTTCC 193

Db     305 GCCCCAATGGCGCACGCGGCGGCGGGTCGCTGGTCGGCGGCGTCAAGTCCGCCTGCTTCC 364

Qy     194 CTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGA 253

Db     365 CGGTGGCAATCGCGCTTGGAGCGACGTGGGACCCGGAGCTCATCGAGCGCGCCGGCGTGG 424

Qy     254 TGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACA 313

Db     425 CGCTGGGAGGACAAGCCAAGAGCAAGGGCGCGTCGGTGCTTCTGGCGCCGACCGTCAACA 484

Qy     314 TGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGG 373

Db     485 TTCACCGCTCCGGCCTCAATGGCCGCAACTTCGAATGCTATTCGGAAGACCCGGCGCTGA 544

Qy     374 CGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGA 433

Db     545 CCGCCGCCTGCGCCGTCGCCTATATCAATGGCGTGCAGAGCCAGGGTGTGGCCGCCACGA 604
```

```
Qy    434 TCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCA 493
Db    605 TCAAGCACTTCGTCGCCAACGAGTCCGAGATCGAGCGGCAGACCATGTCCTCCGATGTCG 664

Qy    494 CGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCC 553
Db    665 ATGAGCGGACGCTGCGCGAAATCTATCTGCCGCCTTTCGAGGAGGCGGTGAAGAAGGCCG 724

Qy    554 AGCCGGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACC 613
Db    725 GCGTGAAGGCCGTCATGTCCTCCTACAACAAGCTCAACGGCACCTATACGAGCGAAAATC 784

Qy    614 CTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCG 673
Db    785 CCTGGCTGCTGACGAAAGTCCTGCGCGAGGAATGGGGCTTCGACGGCGTGGTCATGTCCG 844

Qy    674 ACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGC 733
Db    845 ACTGGTTCGGCTCGCACTCGACGGCTGAAACCATCAATGCCGGGCTCGATCTGGAGATGC 904

Qy    734 CCGGACCTCCACGCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCT 793
Db    905 CGGGGCCTTGGCGGGATCGCGGCGAAAAGCTGGTCGCCGCAGTCCGGGAAGGCAAGGTAA 964

Qy    794 TTATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCTG 853
Db    965 AGGCCGAGACCGTGCGCGCTTCGGCACGGCGTATTCTCCTTCTGCTCGAACGCGTCGGCG 1024

Qy    854 CCTCCGGAGTGACGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCGAAACGGCAG 913
Db   1025 CCTTTGAAAAGGCGCCTGATCTCGCCGAACACGCGCTTGATCTGCCGGAAGATCGTGCGC 1084

Qy    914 CTCTCCTCCGGAAGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGAACAACGTTC 973
Db   1085 TCATCCGCCAACTCGGTGCGGAGGGTGCTGTACTCCTGAAGAATGACGGAGTGCTGCCGC 1144

Qy    974 TGCCCTTGAGCAAGAAGAAGAAGACGCTGATTGTCGGCCCCAACGCCAAGCAGGCCACAT 1033
Db   1145 TTGCCAAGTCGTCCTTCGACCAGATCGCCGTCATCGGCCCCAATGCGGCTTCCGCACGCG 1204

Qy   1034 ACCACGGCGGAGGCTCTGCCGCACTCAGGGCCTACTACGCAGTCACTCCCTTTGACGGCC 1093
Db   1205 TCATGGGCGGAGGAAGCGCGCGGATTGCCGCGCATTATACGGTGAGCCCACTTGAGGGCA 1264

Qy   1094 TCAGCAAGCAGCTCGAGACGCCGCCATCGTACACCGTCGGCGCCTACACCACCGTTCCTC 1153
Db   1265 TTCGCGCGGCGCT--------------------GTCCAACGCCAACAGCCTCCGCCATG 1303

Qy   1154 CCATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGGAGGGTCT 1213
Db   1304 CGGTCGGCTGTAACAACAACCGGCTCATCGACGTCTTCAGCGGCGAGATGACGGTGGAAT 1363

Qy   1214 TCAACGAGCCCCCTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGA 1273
Db   1364 ACTTCAAGGGACGCGGCTTCGAGAGCCGTCCGGTCCATGTCGAGACCGTCGAAAAGGGCG 1423

Qy   1274 CGGACATGCACCTGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTACGCCGACA 1333
Db   1424 AATTCTTCTGGTTCGATCTTCCGTCCGGCGACCTTGATCTCGCCGATTTTTCGGCGCGCA 1483

Qy   1334 TGGAGGGCACGTACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTCGTCTGCG 1393
Db   1484 TGACGGCGACCTTCGTGCCGCAGGAAACCGGTGAACACATCTTCGGCATGACCAATGCTG 1543

Qy   1394 GCACGGCAAAGGCGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCC 1453
Db   1544 GGCTTGCTCGGCTGTTCGTGGACGGCGAACTGGTGGTCGATGGCTATGACGGTTGGACGA 1603

Qy   1454 CCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAATCTCGTCA 1513
Db   1604 AGGGTGAGAACTTTTTTGGAACCGCGAACAGCGAGCAGCGTCGGGCGGTAACGCTTGGGG 1663

Qy   1514 AGGGCAACACGTACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACCCTCAAGG 1573

Db   1664 CCGCACGCCGCTACCGGGTTGTGGTCGAATATG----------AGGCGCCGAAGG 1708

Qy   1574 GCGACACCATCGTCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACG 1633
Db   1709 CCAGCCTGGACGGCATCAACATATGTGCGCTCCGCTTCGGTGTCGAAAAGCCGCT---CG 1765

Qy   1634 ACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCT 1693
Db   1766 GCGATGCCGGGATTGCGGAGGCGGTCGAAACCGCCCGCAAGTCCGATATCGTACTGCTCC 1825

Qy   1694 GCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCC 1753
Db   1826 TCGTCGGCCGTGAGGGCGAGTGGGACACCGAAGGTCTGGATCTGCCCGACATGCGCCTGC 1885

Qy   1754 CCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCG 1813
Db   1886 CGGGTCGCCAGGAGGAGCTGATCGAGGCGGTCGCCGAAACCAATCCCAACGTGGTCGTGG 1945

Qy   1814 TCATGCAGACGGGCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCC 1873
Db   1946 TACTGCAAACGGGTGGTCCCATCGAGATGCCATGGCTCGGCAAGGTGCGTGCGGTGCTGC 2005

Qy   1874 AGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACT 1933
Db   2006 AGATGTGGTATCCCGGCCAGGAACTTGGCAATGCGCTTGCGGACGTTCTCTTTGGTGATG 2065

Qy   1934 ACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAACCCCGCGT 1993
Db   2066 TCGAGCCTGCCGGCCGCTTGCCACAGACCTTCCCGAAGGCGCTCACGGATAATTCCGCCA 2125

Qy   1994 TTCTCAACTTCCGCAC----------CGAGGCCGGGCGCACGCTGTACGGCGAGGACG 2041
Db   2126 TTACCGACGATCCGTCGATCTATCCTGGCCAGGACGGCCATGTGCGCTACGCGGAAGGGA 2185

Qy   2042 TCTACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCC 2101
Db   2186 TCTTCGTCGGCTATCGCCATCACGATACAAGAGAGATCGAACCACTCTTCCCCTTCGGCT 2245

Qy   2102 ACGGCCTGTCCTACACCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGA----- 2156
Db   2246 TCGGTCTTGGCTACACCCGCTTTACCTGGGGTGCCCCGCAACTATCGGGAACGGAAATGG 2305

Qy   2157 ----CGGCAAGCTGAGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCAC 2212
Db   2306 GGGCGGATGGTCTTACGGTGACGGTCGATGTCACCAATATAGGCGACAGGGCGGGATCGG 2365

Qy   2213 AGGTGGCCCAGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGG 2272
Db   2366 ACGTGGTGCAGCTCTATGTTCACTCTCCCAA---TGCCAGGGTCGAGCGGCCGTTCAAGG 2422

Qy   2273 AGCTCAAGGGCTTCGCAAAGGTCGAACTGCAGCCCGGCGAGACGAAGGCCGTGACAATCG 2332
Db   2423 AGCTGCGTGCCTTTGCGAAGCTCAAGCTGGCCCCGGGCGCGACCGGTACGG---CGGTGC 2479

Qy   2333 AGGAGCAGGAGAAGTACGTCGCTGCGTATTTTGATGAGGAGCGGGATCAGTGGTGTGTCG 2392
Db   2480 TGAAGATCGCTCCTCGCGACTTGGCTTACTTCGATGTCGAGGCCGGTCGTTTCCGGGCTG 2539

Qy   2393 AAAAGGGTGACTATGAGGTTATCGTGAGCGACAGC 2427
Db   2540 ATGCGGGCAAGTACGAGCTGATCGTGGCGGCCAGC 2574
```

RESULT 9
ABI293760
LOCUS ABI293760 2636 bp mRNA linear PLN 09-AUG-2000
DEFINITION Agaricus bisporus mRNA for putative beta glucosidase (bg1 gene), strain D649.
ACCESSION AJ293760
VERSION AJ293760.1 GI:9796011
KEYWORDS beta glucosidase; bg1 gene.
SOURCE Agaricus bisporus.

```
  ORGANISM  Agaricus bisporus
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Agaricaceae; Agaricus.
REFERENCE   1  (bases 1 to 2636)
  AUTHORS   Morales-Almora,P. and Thurston,C.F.
  TITLE     Molecular analysis of the cellulolytic genes in Agaricus bisporus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2636)
  AUTHORS   Morales-Almora,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2000) Morales-Almora P., Microbiology, King's
            College London, 150 Stamford Street, London, SE1 8WA, UNITED
            KINGDOM
FEATURES             Location/Qualifiers
     source          1. .2636
                     /organism="Agaricus bisporus"
                     /strain="D649"
                     /db_xref="taxon:5341"
                     /clone="PMRT1615"
                     /country="United Kingdom"
     gene            25. .2610
                     /gene="bg1"
     CDS             25. .2610
                     /gene="bg1"
                     /EC_number="3.2.1.21"
                     /function="putative hydrolysis of beta-D-glucose"
                     /codon_start=1
                     /product="putative beta glucosidase"
                     /protein_id="CAC03462.1"
                     /db_xref="GI:9796012"
                     /translation="MPPSDFANADIDKIVDALTTDEAILLSAGVGFWHTHAIERLQIP
                     AVKVSDGPNGIRGNHFFMGTPAKCLPSSTAMGATFDRDLLEEVGLKLLAREAKLRSAS
                     VILAPTCNIQRNPLGGRSFESFSEDPFLSGMLCSAYINGIQKGGIGATIKHFVGNDKE
                     DDRTGYDSIHYRTTSSGDLSPSVHAGSRNMPLHGAIMTAYNRVNGIHVCENPELLQKI
                     LRDEWKSDVMIMSDWFGTYSVDVGLNAGLDLEMPGLNKWRSLESVNRSIQSRKVTAKK
                     VKERARKVLELVKKCAQGAPEILDGDGLERTLDSDEDNALMRKFAAESIVLLKNDKKV
                     LPLDRKSLKKVAIVGGNAKAAVLSGGGSAALKPSYFVARPAGLVKALKEVNPNIEVTY
                     SEGARAAKTLPTLDFDLFTESGQRGWTGSWHAHENDDSLVALPTPIKTQYVDETRMFI
                     SSSSPAGITKKWTLRMKGYLKPRERDCKFEFGLTAAGRAKLFVDGKLVIDNWTRQRRG
                     VEFFGCGSEEETGVVELKAGVKHEIYVDFCNVRGPADGDETETIMDSNPGIRLGGAEV
                     EDPDELLNSAVSIAKDADAVIIVVGLNGDWETEGNDRTTLALPGRTDELVQKVVAANP
                     RAVVVTQAGSAITMPWVDQVSSLLHTWYLGNATGDAIADVIFGNHNPSGKLSLTFPKR
                     LEDVPAHGYFHSEKGKVTYAEGLYVGYKHYQHRNIEPLFAFGHGLSYTSFNISDLRLS
                     QPIVSGDQFDLKATVKLTNTGNITGSQVVQLYIGLPKTSELTHPRWQLRGFKKMRDVK
                     PGESREVELVMDRLSVSYWDKEWVVENGAYDVRVAFTSEEGVGEGQELLGRFNVEKGF
                     GWRGL"
BASE COUNT      657 a    603 c    682 g    694 t
ORIGIN

  Query Match           12.1%;  Score 302.2;  DB 8;  Length 2636;
  Best Local Similarity  49.8%;  Pred. No. 1.6e-27;
  Matches 1100;  Conservative    0;  Mismatches 1048;  Indels   59;  Gaps   11;

Qy       2 TGGCTGATATTGATGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGATC 61
Db      41 TTGCCAACGCTGACATTGATAAAATTGTCGACGCATTGACTACCGATGAAGCCATTCTTT 100

Qy      62 TGCTGGCTGGTATCGACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCTC 121
Db     101 TGAGTGCCGGAGTTGGATTTTGGCATACTCATGCAATTGAGAGGCTGCAGATTCCCGCTG 160

Qy     122 TCCGCTTTACAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCTTCAATGGCGTCCCTG 181
Db     161 TGAAGGTTAGCGACGGGCCCAATGGTATCCGAGGAAATCATTTCTTTATGGGCACGCCAG 220

Qy     182 CGGCCTGCTTCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAG 241
Db     221 CCAAGTGTTTGCCGTCATCCACTGCAATGGGCGCAACTTTCGATCGTGACCTACTTGAAG 280

Qy     242 AGGCAGGT---AAGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCG 298
Db     281 AAGTTGGTCTCAAGCTTCTCGCTAGAGAAGCAAAGCTCAGATCCGCCTCTGTTATCCTCG 340

Qy     299 GCCCGACTATCAACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTG 358
Db     341 CTCCAACTTGTAACATCCAACGTAATCCCCTCGGTGGTAGGAGTTTTGAAAGTTTTTCGG 400

Qy     359 AGGATCCGTTCCTGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGCACTG 418
Db     401 AGGACCCTTTCCTGTCTGGTATGCTTTGCAGTGCCTACATCAACGGAATTCAGAAAGGGG 460

Qy     419 GAGTGCAGGCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCA-TGATG 477
Db     461 GTATTGGCGCCACTATCAAGCATTTCGTGGGAAATGACAAGGAAGATGATCGAACTGGAT 520

Qy     478 GTGCAGAGCATCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATT 537
Db     521 ATGACAGTATTCATTACCGAACGACCTCTTCGGGAGATTTATCTCCTTCCGTTCATGCTG 580

Qy     538 GCTGTGCGAGACTCCCAGCC--GGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCG 595
Db     581 GCTCAAGAAATATGCCTCTCCATGGGGCGATCATGACCGCGTACAACCGTGTTAACGGTA 640

Qy     596 TGTCGTGCAGCGAGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGG 655
Db     641 TTCACGTTTGCGAGAATCCCGAACTTCTACAGAAAATATTACGCGATGAGTGGAAATCAG 700

Qy     656 ATGGCCTAATCATGAGCGACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAG 715
Db     701 ATGTCATGATCATGAGCGACTGGTTCGGCACGTACAGCGTCGATGTGGGTTTAAATGCCG 760

Qy     716 GCCTCGACCTCGAGATGCCCGGACCTCCACGCTTCCG---AGGAGAAACACTCAAGTTCA 772
Db     761 GTCTTGACCTTGAAATGCCCGGCCTTAACAAATGGCGTTCACTCGAGTCTGTCAACCGAT 820

Qy     773 ACGTCTCCAACGGAAAGCCCTTTATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTC 832
Db     821 CAATTCAATCTAGGAAAGTAACCGCCAAGAAAGTCAAAGAGCGTGCTCGAAAAGTCCTCG 880

Qy     833 AGTTCGTCAAGAAGTGTGCT------GCCTCCGGAGTGACGGAGAACGGCCCCGAGACGA 886
Db     881 AGCTTGTCAAAAAATGTGCTCAAGGCGCCCCTGAAATTTTGGACGGCGATGGCCTTGAAC 940

Qy     887 CTGTCAACAACACCCCCGAAACGGCAGCTCTCCTCCGGAAGGTTGGCAACGAGGGCATCG 946
Db     941 GCACTCTTGATTCAGACGAAGATAATGCTCTCATGCGGAAATTTGCTGCAGAGTCTATTG 1000

Qy     947 TGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCAAGAAG------AAGAAGACGC 1000
Db    1001 TTCTGCTCAAGAACGACAAGAAAGTCCTTCCCTTGGACCGTAAGAGTTTAAAGAAAGTTG 1060

Qy    1001 TGATTGTCGGCCCCAACGCCAAGCAGGCCACATACCACGGCGGAGGCTCTGCCGCACTCA 1060
Db    1061 CAATCGTGGGTGGTAATGCCAAAGCCGCTGTACTATCTGGTGGAGGTTCTGCTGCCCTCA 1120

Qy    1061 GGGCCTACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGCTCGAGACGCCGCCAT 1120
Db    1121 AGCCGTCCTACTTCGTCGCGCGGCCGGCCGGGCTGGTCAAGGCTCTGAAAGAAGTCAACC 1180

Qy    1121 CGTACACCGTCGGCGCCTACACC----------------------------ACCGTTCCTCC 1154
Db    1181 CAAACATCGAGGTTACTTACAGCGAAGGCGCTAGAGCGGCCAAGACGTTACCGACTTTGG 1240

Qy    1155 CATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTCGAGGGTCTT 1214
Db    1241 ATTTCGATTTGTTCACTGAGTC-CGGCCAACGTGGCTGGACTGGGTCTTGGCATGCTCAT 1299

Qy    1215 CAACGAGCC----CCCTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCA 1270
Db    1300 GAAAATGACGATAGCCTCGTTGCCTTGCCTACGCCGATAAAAACACAATATGTTGACGAG 1359

Qy    1271 AGACGGACATGCACCTGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTACGCCG 1330
Db    1360 ACAAGAATGTTCATCAGCAGTTCTTCTCCTGCCGGTATAACCAAGAAGTGGACGCTTCGT 1419

Qy    1331 ACATGGAGGGCACGTACACCGCCGACGAG-GACTGCACCTACGAGCTCGGCCTCGTCGTC 1389
```

```
Db   1420 ATGAAAGGTTATTTGAAACCGCGAGAGCGTGATTGTAAATTCGAATTTGGATTGACCGCT 1479

Qy   1390 TGCGGCACGGCAAAGGCGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAG 1449

Db   1480 GCTGGAAGAGCCAAGCTGTTCGTCGATGGAAAACTGSTGATTGACAACTGGACGCGTCAA 1539

Qy   1450 GTCCCCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAATCTC 1509

Db   1540 CGTCGCGGTGTAGAATTCTTTGGATGTGGTTCAGAAGAGGAGACTGGCGTTGTTGAGTTG 1599

Qy   1510 GTCAAGGGCAACACGTACAAGTTCAAGATCGAGTTC------GGCTCCGCACCCACCTAC 1563

Db   1600 AAAGCCGGTGTCAAGCATGAGATTTATGTCGACTTCTGTAATGTTCGTGGGCCAGCTGAT 1659

Qy   1564 ACCCTCAAGGGCGACACCATCGTCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAG 1623

Db   1660 GGTGATGAAACCGAAACAATCATGGATAGTAACCCTGGTATTCGCCTTGGGGGCGCTGAA 1719

Qy   1624 GTCATTGACGACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAG 1683

Db   1720 GTTGAAGATCCTGATGAGCTCCTCAATTCCGCTGTCTCAATCGCAAAAGACGCTGATGCC 1779

Qy   1684 GTCATCATCTGCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGC 1743

Db   1780 GTTATTATCGTTGTCGGTTTAAACGGTGATTGGGAGACAGAGGGTAATGACCGTACCACC 1839

Qy   1744 ATGAAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAAC 1803

Db   1840 CTTGCCCTTCCCGGCCGTACCGACGAATTGGTACAAAAAGTCGTTGCTGCAAATCCTCGT 1899

Qy   1804 ACCGTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCC 1863

Db   1900 GCTGTCGTCGTCACTCAAGCGGGTTCCGCTATAACTATGCCTTGGGTCGACCAAGTTTCT 1959

Qy   1864 GCCGTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTC 1923

Db   1960 TCCCTCCTACATACCTGGTATCTAGGTAATGCTACTGGCGATGCTATCGCCGACGTCATT 2019

Qy   1924 TTTGGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGAC 1983

Db   2020 TTTGGGAATCATAATCCTTCTGGAAAGCTTTCATTGACCTTCCCGAAGCGATTGGAGGAT 2079

Qy   1984 AACCCCGCGTTTCTCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTC 2043

Db   2080 GTTCCTGCTCATGGCTATTTCCATTCTGAGAAAGGAAAGGTGACGTACGCTGAAGGTCTT 2139

Qy   2044 TACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCAC 2103

Db   2140 TACGTGGGTTATAAACATTATCAACATCGTAATATCGAACCTCTCTTCGCGTTTGGCCAT 2199

Qy   2104 GGCCTGTCCTACACCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCA 2150

Db   2200 GGTTTATCGTACACGTCGTTCAATATTTCAGACCTTCGTCTCTCACA 2246


RESULT 10
REU80928_2/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS REU80928 Accession U80928
     Fragment Name      Begin     End
     REU80928_0             1    110000
     REU80928_1        100001    210000
     REU80928_2        200001    310000
     REU80928_3        300001    371255
Continuation (3 of 4) of REU80928 from base 200001 (U80928 Rhizobium etli strain CFN42 s

  Query Match            10.8%;  Score 270;  DB 1;  Length 110000;
  Best Local Similarity  48.0%;  Pred. No. 4.5e-24;
  Matches 1116;  Conservative    0;  Mismatches 1135;  Indels   75;  Gaps    9;

Qy     10 ATTGATGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGATCTGCTGGCT 69

Db   81149 ATTGATAAGAAAGCATTGCTCGATAATATGACCTTGGCGGAGCAGGTCTCGCTACTTTCC 81090

Qy     70 GGTATCGACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCTCTCCGCTTT 129

Db   81089 GGCGATACGTTCTGGTCTTTACCGCCCATCGACCGCCTAGGGATAGGGAGATTGCGCTTG 81030

Qy    130 ACAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCT---TCAATGGCGTCCCTGCGGCC 186

Db   81029 ACCGACGGCCCCAACGGCGCCCGCGGGGCGGGGTCTTTTGTCGGAGGTGTGACTGCAGCG 80970

Qy    187 TGCTTCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCA 246

Db   80969 GCTTTCCCGGTGGGCATCGCCATCGGAGCGAGTTGGAACCCGGATCTAGCCAAGGAAATC 80910

Qy    247 GGTAAGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACT 306

Db   80909 GGCAGTGCCCTAGGTGACGAGGTTCTTTCAAAAGGTGCCCACGTCTCGTTAGCCCCCACC 80850

Qy    307 ATCAACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCG 366

Db   80849 GTTAACATCCAGCGTAGCGTCACAAACGGCCGCAACTTTGAGTGCTTCTCTGAGGATCCG 80790

Qy    367 TTCCTGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAG 426

Db   80789 ATCCTGACGGCGGAACTTGCAGTCGGCTATATCGAAGGTCTGCAGTCCACGAAAGTCGGT 80730

Qy    427 GCTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGC 486

Db   80729 GCCACGATAAAACATTTCGTCGGCAACGAGTCCGAGATTGAACGTACAACCATCTCTTCA 80670

Qy    487 ATCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGA 546

Db   80669 GATATCGATGAACGCACGCTGCGCGAAGTCTACCTGATACCCTTTGAGACGGCGGTGAAG 80610

Qy    547 GACTCCCAGCCGGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGC 606

Db   80609 CGGGCAAAGGTCTGGGCCGTGATGTCTTCGTATAACAAGCTAAACGGTACTTACACGGCG 80550

Qy    607 GAGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATC 666

Db   80549 GAAAGCCATTGGCTGTTGAACGAGGTTCTGCGCGGTGACTGGGGTTTTAACGGCGTGGTG 80490

Qy    667 ATGAGCGACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTC 726

Db   80489 ATGTCGGACTGGTTCGGCTCGCGTTCGACCGCACCCACCGTGAATGCCGGGCTGGATCTG 80430

Qy    727 GAGATGCCCGGACCTCCACGCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGA 786

Db   80429 GAGATGCCGGGCCCGACTCGCGATCGCGGCTCCAAGCTACTGGCTGCGGTCGAAGGCGGC 80370

Qy    787 AAGCCCTTTATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAG 846

Db   80369 GAGGTTAGTGTCGAGACGATCCGCGCCTGTGTGCGCAATATCCTGACTTTGATGGAACGC 80310

Qy    847 TGTGCTGCCTCCGGAGTGACGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCGAA 906

Db   80309 ACCGGCGCTATCAACGATCATCGCGAGTTTAAGGAGTACGCCATTGATCAACCGAAACAT 80250

Qy    907 ACGGCAGCTCTCCTCCGGAAGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGAAC 966

Db   80249 AGGGCATTGATCAGGCGCGCGGGAGCGGAAAGCGCAGTTCTGCTACAGAATGACGGGATC 80190

Qy    967 AACGTTCTGCCCTTGAGCAAGAAGAAGAAGACGCTGATTGTCGGCCCCAACGCCAAGCAG 1026

Db   80189 ------CTGCCATTGGCTCAGCAAGGTATGGTCGCCATCATTGGACCCAATGCCAAGGTC 80136

Qy   1027 GCCACATACCACGGCGGAGGCTCTGCCGCACTCAGGGCCTACTACGCAGTCACTCCCTTT 1086

Db   80135 GCGCAGGTGATGGGAGGCGGCTCGGCGCAGTTGAACCCTCACTATGTCATCAGTCCGTGG 80076

Qy   1087 GACGGCCTCAGCAAGCAGCTCGAGACGCCGCCATCGTACACCGTCGGCGCCTACACCACC 1146

Db   80075 CAGGGGCTGGTGGACGCGCTGGGTGAGGAGAATCTGTGCTACGCTCAGGGCTGCAACAAT 80016
```

```
Qy   1147 GTTCCTCCCATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGG 1206
Db  80015 TATC-------GGTTTCAGCCGCTAATCGAAAACCCCACAACCTTCGAGTTTTTCCAAGG 79963

Qy   1207 AGGGTCTTCAACGAGCCCCCTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTC 1266
Db  79962 AAGGGAGCTAGCTGGCGAACCGGT-----------GAAGGTCGTCGAGGAACCGTCGAGC 79914

Qy   1267 ACCAAGACGGACATGCACCTGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTAC 1326
Db  79913 CTTGGTGTATGGTTGCCCCCCGTGGCCGAGGGCTTTGTCGATCCCCTTCGC---TTTTCG 79857

Qy   1327 GCCGACATGGAGGGCACGTACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTC 1386
Db  79856 GCTAGAATGCGCACCATCTTCACAGCCTCGGAAGCCGGCGTCTATCGCGTCGGACTAACC 79797

Qy   1387 GTCTGCGGCACGGCAAAGGCGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAG 1446
Db  79796 TCGGCCGGGCTTGGCCGGGTCTATGTGGACGGCAGACTTGTGGTGGACGCTTGGGCTTCC 79737

Qy   1447 CAGGTCCCCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAAT 1506
Db  79736 TGGATACGTGGTACCACATTCTTTGAAGAAGGCTGCGAAGAGGTCGTGGGTGAAATCACG 79677

Qy   1507 CTCGTCAAGGGCAACACGTACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACC 1566
Db  79676 CTCGAAGCCGGCCGTACATACGAGGTCGTCGCCGAGTACGCCCGGCACGACCACGTCAAC 79617

Qy   1567 CTCAAGGGCGACACCATCGTCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTC 1626
Db  79616 CTCTATATCGCCGCAATTCGGGTAGGAATAGGC----------------------AGA 79581

Qy   1627 ATTGACGACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTC 1686
Db  79580 TTTTCGGCCGAAGCGGAGATTGCCGAGGCAGCAGCCGTCGCGGCAAAGGCTGATCATGCG 79521

Qy   1687 ATCATCTGCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATG 1746
Db  79520 GTAGTCTTCGTGGGCAGAACGGGCGACTGGGATACCGAGGGCTCCGATCTTCGCGGTATT 79461

Qy   1747 AAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACC 1806
Db  79460 GCACTTCCCGGTCTGCAGAACCAGCTTGTTGAGGCGGTAATTGCGGCCAATCCGAACACG 79401

Qy   1807 GTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCC 1866
Db  79400 ATCGTCGTACTACAAACCGGTGGACCGGTGGAAATGCCCTGGCTTTCCGGTGCTCGTGCA 79341

Qy   1867 GTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTT 1926
Db  79340 GTACTGCAATGCTGGTATCCCGGACAGGAGGCTGGCAACGCGATCGCCGATGTGCTCCTC 79281

Qy   1927 GGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAAC 1986
Db  79280 GGCAAGGCCGAGCCCTCGGGCCGACTGGCACAGACCTTCCCCGTGCGCTGGGCTGACAAC 79221

Qy   1987 CCCGCGTTTCTCAACTTCCGCACCG----------AGGCCGGGCGCACGCTGTACGGC 2034
Db  79220 CCCACACATACTGAGGATGACGCCGTCTATCCAGGTAAGGACGGCCATGTGCGATATGAC 79161

Qy   2035 GAGGACGTCTACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCC 2094
Db  79160 GAAGGTGTGTTCGTCGGTTATCGGCACTACGACCGCCACGGCATTAAGCCGCTGTTCCCG 79101

Qy   2095 TTTGGCCACGGCCTGTCCTACACCACTTTTGCCTTTTCCAATCTCTCCG------TGTCT 2148
Db  79100 TTCGGTCATGGTCTCGGTTATTCAAGCTTTGCGATGTCGGACCTGACGGTAGGGCTACCA 79041

Qy   2149 CACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGC 2208
Db  79040 GACGCTGCCGGCGCGGTGACAGTGACACTGGAATTGACCAACATCAGCGAGCGGCCAGGT 78981

Qy   2209 GCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTC 2268
Db  78980 TCGGCAGTGGTGCAAATCTATGTCGGCGATGTTGAAGCATC---TATACCCAGACCGGTT 78924

Qy   2269 AAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCAGCCCGGCGAGA 2314
Db  78923 AAAGAGCTGAAAGCCTTCTCGAAAATTGCCCTGGAACCCGGTGAGA 78878
```

RESULT 11
AX489445
LOCUS AX489445 2601 bp DNA linear PAT 16-AUG-2002
DEFINITION Sequence 6745 from Patent WO02053728.
ACCESSION AX489445
VERSION AX489445.1 GI:22323457
KEYWORDS .
SOURCE Candida albicans.
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1
AUTHORS Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Patent: WO 02053728-A 6745 11-JUL-2002;
Elitra Pharmaceuticals, Inc. (US)
FEATURES Location/Qualifiers
source 1. .2601
/organism="Candida albicans"
/db_xref="taxon:5476"
BASE COUNT 917 a 334 c 484 g 866 t
ORIGIN

Query Match 8.2%; Score 205; DB 6; Length 2601;
Best Local Similarity 46.7%; Pred. No. 7.6e-16;
Matches 1031; Conservative 0; Mismatches 1097; Indels 78; Gaps 9;

```
Qy      11 TTGATGTTGAGGCCATCTTGAAGAAGCTCACCCTGGCCGAGAAGGTCGATCTGCTGGCTG 70
Db      14 TTGATATTGATAATATATTATCTCAATTGACATTAGAAGAAAAAATTGGATTAGTAGGTG 73

Qy      71 GTATCGACTTCTGGCACACAAAGGCTCTCCCCAAGCATGGAGTCCCCTCTCTCCGCTTTA 130
Db      74 GAATTGATTTTTGGCATACATATCCAATTTCTCGATTAAACATTCCCAAAGTGAGATTCA 133

Qy     131 CAGATGGCCCCAACGGCGTAAGAGGGACCAAGTTCTTCAATGGCGTCCCTGCGGCCTGCT 190
Db     134 CTGATGGTCCCAATGGTATTAGAGGCACAAGGTTTTTCAATGGTGTTCCTCTGGCTTGTT 193

Qy     191 TCCCTTGCGGCACGTCGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTA 250
Db     194 TCCCCTGTGGAACTGGATTAGCTGCTACTTTTGATCATGAATTATTATTGACAACCGGGA 253

Qy     251 AGATGATGGGCAAAGAGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCA 310
Db     254 AATTAATGAACATTGAAGCTAAATTTAAAAATGCGCACGTGATCTTGGGTCCTACGATGA 313

Qy     311 ACATGCAACGCTCCCCTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCC 370
Db     314 ATATTCAACGAGGCCCATTAGGTGGCCGAGGATTTGAATCATTTTCTGAAGATCCTTATT 373

Qy     371 TGGCGGGCTTGGGAGCTGCGGCTCTCATCCGCGGCATTCAGAGC---ACTGGAGTGCAGG 427
Db     374 TAACTGGACAAATTGCCAGTGCTATAATTAAGGGTATTCAATATGATAATGAAATTGGTG 433

Qy     428 CTACGATCAAGCACTTTTTGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCA 487
Db     434 CCACGGTTAAACATTATGTTTGTAATGATTTAGAAGATGAAAGGAGTGCTAGTGATTCTT 493

Qy     488 TCGTCACGGAGCGGGCTCTCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAG 547
Db     494 TGGTTACACCAAGAGCATTGAGAGAAATTTATTTAGAACCATTTAGAATAGCTATTAAAG 553

Qy     548 ACTCCCAGCCGGGTGCGTTCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCG 607
```

```
Db    554 AAAGTAATCCAATTTGTTTGATGACAGGGTATAATAAAGTTAATGGTGAACATGTTTCTC 613

Qy    608 AGAACCCTAAATATCTTGATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCA 667

Db    614 AAAGTAAATTTTTCTTGCAAAATATATTACGTGATGAATGGAATTGGCAAGGGACAATTA 673

Qy    668 TGAGCGACTGGTACGGCACATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCG 727

Db    674 TATCTGATTGGTATGGGACATATACAAGTAAAAAAGCCATTGAAAATGGTTTGGATTTAG 733

Qy    728 AGATGCCCGGACCTCCACGCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAA 787

Db    734 AAATGCCAGGATCTCCAATTTTCCGTAATAAACAATTGTTGACTAGTATGATTAAATCAA 793

Qy    788 AGCCCTT---TATCCACGTCATTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAG- 843

Db    794 AAGAATTGCATATCAAACATTTGGATGATCGTGTAAAGAATGTTTTAAAATTAATTAAAT 853

Qy    844 -----AAGTGTGCTGCCTCCGGAGTGACGGAGAACGGCCCCGAGACGACTGTCAACAACA 898

Db    854 TTGCCAAACAGTCAAGTGTTGTGGTCACTGAAGATGGTAAAGAATCATCGGAGAATAACA 913

Qy    899 CCCCCGAAACGGCAGCTCTCCTCCGGAAGGTTGGCAACGAGGGCATCGTGCTGCTGAAGA 958

Db    914 CTCAAGAAACAAGAGATATACTACGCAAACTAGCTCAAGATTCAATTGTTTTATTGAAAA 973

Qy    959 ACGAGAACAACGTTCTGCCCTTGAGCAAGAAG----------------AAGAAGACGC 1000

Db    974 ATGATAATAATTTATTACCATTGAAAAGGGATGACGTCGACTTCAGCTCCAAGTCTATTG 1033

Qy   1001 TGATTGTCGGCCCCAACGCCAAGCAGGCCACATACCACGGCGGAGGCTCTGCCGCACTCA 1060

Db   1034 CAATAATTGGTCCTAATGCAAAAATTGCCGCTTATTCAGGTGGTGGGTCAGCCACTTTAC 1093

Qy   1061 GGGCCTACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGCTCGAGACGCCGCCAT 1120

Db   1094 CAGCTTATTACACCACTACTCCTTATAATGCCATTGTTGAGAAATTATCTTCTATTTCCA 1153

Qy   1121 CGTACACCGTCGGCGCCTACACCACCGTTCCT----------------------------C 1153

Db   1154 AATTTGATATTACATCACAACTCAAATATACTATTGGTGCAAAAGCATATAAATATTTAC 1213

Qy   1154 CCATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGGAGGGTCT 1213

Db   1214 CAGAATTAGGTCCACAAGTGATCAATCCCAAAACTGGTAAACCCGGGTTTTCCATGAAAT 1273

Qy   1214 TCAA---------CGAGCCCCCTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCT 1264

Db   1274 TTTATAAAAAGCCAAAATCCGTTCCTAATGAAAATCGAGAATTATTTGATGAATTAGATA 1333

Qy   1265 TCACCAAGACGGACATGCACCTGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGG- 1323

Db   1334 CAGAAATATCTGATATTTTATTAGGTGATTATTATCATAAAGATATCCCTTCTAATGGGT 1393

Qy   1324 -----TACGCCGACATGGAGGGCACGTACACCGCCGACGAGGACTGCACCTACGAGCTCG 1378

Db   1394 TATATTATATTGATTTTGAATGTGAATTCACACCATCAAAAACCCAACATTATGAATTTG 1453

Qy   1379 GCCTCGTCGTCTGCGGCACGGCAAAGGCGTACGTAGACGACCAGCTCGTCGTCGACAACG 1438

Db   1454 GTTTAACTGTTCATGGTACAGCACAATTATTTATTGATGATAAACTTGTTGTTGATAATA 1513

Qy   1439 CCACCAAGCAGGTCCCCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAGGAGACGGGCC 1498

Db   1514 AAACCAAACAAATTAAAGGAGTTTCATTTTTGAATTCTGGAACCATTGAAGAACGAGGAT 1573

Qy   1499 GCATCAATCTCGTCAAGGGCAACACGTACAAGTTCAAGATCGAGTTCGGCTCCGCACCCA 1558

Db   1574 CAATTGAATTACATCAAGGTAAAACTTATAAAATCATTGTTGAATATGGTTCGGCACCTA 1633

Qy   1559 CCTACACCCTCAAGGGCGACACCATCGTCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCT 1618

Db   1634 CATTTACATTAAAAGATCAAGTTGGAGAATATTTTGGTGGTGGGATTCGTCTTGGTATGA 1693

Qy   1619 GCAAGGTCATTGACGACCAGGCC---GAAATCGAAAAGTCCGTCGCCCTCGCCAAGGAGC 1675

Db   1694 ATGAAATTATTAATGATGATGAACAAGAAATCATCAATGCAGTAAATTTGGCTAAATCAG 1753

Qy   1676 ACGACCAGGTCATCATCTGCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGCGCCGACC 1735

Db   1754 TAGATTTAGTGATTTTAATTATTGGATTAAATAAAGATTGGGAAAGTGAATCTTATGATC 1813

Qy   1736 GCGCGAGCATGAAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCCGCCGCGA 1795

Db   1814 GTCCTGATATGAAATTACCTGGATTACAAGATAAATTGATTGAATCGGTATTGGACGTTA 1873

Qy   1796 ACCCAAACACCGTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATGCC-- CTGGCTCG 1852

Db   1874 ATCCAAACACCATTATTGTTAATCAATCCGGTACACCTGTGGAATTTTCTGAATGGTTGG 1933

Qy   1853 ACGCCACGCCCGCCGTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTG 1912

Db   1934 ATAAACTGAAAGCTTTGTTGCATTGTTGGTATGGTGGGAATGAAAGTGGGAATGCCATTG 1993

Qy   1913 CCGACGTCGTCTTTGGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGC 1972

Db   1994 CTGATATTTTATTTGGTAATGTCAATCCTAATGGGAAATTATCATTGACTTTCCCATTGA 2053

Qy   1973 GCCTGCAGGACAACCCCGCGTTTCTCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACG 2032

Db   2054 AAAATAATGATAATCCAACTTTTTTGAATTTTAAAACTGAAAGAGGTCGAGTATTATATG 2113

Qy   2033 GCGAGGACGTCTACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCC 2092

Db   2114 GAGAAGATATTTTTGTTGGTTATAGATATTATGAAAAATTGAATCGTCAAGTTGCTTTCC 2173

Qy   2093 CCTTTGGCCACGGCCTGTCCTACACCACTTTTGCCTTTTCCAATCT 2138

Db   2174 CATTTGGATTTGGATTATCATATACTAAATTTGAATTTTTTGATTT 2219
```

RESULT 12
SC8A6

```
LOCUS       SC8A6                37445 bp    DNA     linear   BCT 12-MAY-2002
DEFINITION  Streptomyces coelicolor cosmid 8A6.
ACCESSION   AL031013 AL645882
VERSION     AL031013.2  GI:20520810
KEYWORDS    beta-glucosidase; elongation factor G; fusA; sucC; succinyl-coa
            synthetase alpha chain; succinyl-coa synthetase beta chain; sucD;
            TPP-requiring enzyme; transfer-RNA-Leu; TTA Leu codon.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 37445)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
   PUBMED   8843436
REFERENCE   2  (bases 1 to 37445)
  AUTHORS   Seeger,K.J. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 37445)
  AUTHORS   Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     On May 9, 2002 this sequence version replaced gi:3288600.
            Notes:
```

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 8A6 lies between 3F9 and 1F2 on the AseI-A genomic restriction fragment.

FEATURES Location/Qualifiers

source 1. .37445
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 8A6"

misc_feature 2. .107
/note="nominal overlap with St8A6 from 1 to 106
nominal overlap with cosmid 3F9 from 19725 to 19830"

stem_loop 657. .711
/note="hairpin loop with 23 bp stem"

gene complement(730. .2160)
/gene="SCO6581"

CDS complement(730. .2160)
/gene="SCO6581"
/note="SC8A6.02c, probable transmembrane transport protein, len: 476 aa; some similarity to e.g. TR:Q51330 (EMBL:U4007 5) oxalate:formate antiport protein from Oxalobacter formig enes (418 aa), fasta scores; opt: 226 z-score: 230.1 E(): 1 .4e-05, 24.1% identity in 436 aa overlap"
/codon_start=1
/transl_table=11
/product="putative transmembrane transport protein"
/protein_id="CAA19774.1"
/db_xref="GI:3288602"
/db_xref="SPTREMBL:O87836"
/translation="MGQGALARSFDARGAIDMTTTDVTRVAAYREVTDRNGRVYRVGE
SDIDLMGRKRKWMVILPWIGMMGISSAEYAFASAEDTLHEAQHWSSGSIYWMMTAWVF
CQAAVAFPAGRLRESGKLPARWAMMLGSAGTLLGYLSLAFAPHVSLAFIGFGVFSGMG
AGMVYATCVNMVGKWYPERRGGKTGFVNGGFAYGSVPFVFIFHGYMDTSNFRWVLVSA
GVFLAAMVAIAGFYFRDPPKNWWPAAVDPLNPPADPRARRSLEKNPPAVKQYSPMEAW
KTGRVALMWFCLACTSGVNIFGIAFQVDIGEEAGFAAGVVAAAMSLKAIVNGTGRGVI
GWLSDLYGRKQCLLYVCAILGLAQFGIIWSAEIKNLPLFLVFSAISGFGGGAIFPMFA
ALTADYFGENNNATNYGMVYSSKLVSGLGAGMGSVVVGVWGYNGAFSLAGSISIFAGF
VALSLRPPGRPKGKRVTPNPQPLGEG"

gene complement(2233. .4377)
/gene="SCO6582"

CDS complement(2233. .4377)
/gene="SCO6582"
/note="SC8A6.03c, unknown, len: 714 aa; similar to many hypothetical proteins including several from each of the archeons Archaeoglobus fulgidus TR:O28341 (EMBL:AE000970) conserved hypothetical protein AF1938(673 aa), fasta scores; opt: 960 z-score: 1289.8 E(): 0, 35.6% identity in 693 aa overlap and Pyrococcus horikoshii D1028624 (EMBL:AB009518) hypothetical protein PHCV016 (460 aa), fasta scores; opt: 1280 z-score: 1246.9 E(): 0, 44.3% identity in 461 aa overlap. Contains PS00216 Sugar transport proteins signature 1. Also similar to S. coelicolor hypothetical protein SC9B10.09 (E() : 6.1e-15, 33.9% identity in 375 aa overlap)"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAA19775.1"
/db_xref="GI:3288603"
/db_xref="SPTREMBL:O87837"
/translation="MAEDRLARVRGLLDAVRGEGRTALTAPEGKVLADAYGIAVPGEE
LARDVDEAVACAARFGGPVVMKIVSPDILHKTDAGGVVVGVEGAADVRAAFCRIVDNA
RAYDASARIEGVQIQELLPQGQEVIVGAVTDPTFGKVVAFGLGGVLVEVLKDVTFRLA
PVDADEALSMLDSIRAAEVLRGVRGQAGVDRWAVAEQIRRVSELVADFPEIAEVDLNP
VIATPEGAVAADIRVILATDTPKERRRYTREEILTSMRRLMQPASVAVIGASGEPGKI
GNSVMRNLVDGGFAGEIHPVNPKADDILGRKAYKSVTDVPGEVDVAVFAIPARFVAAA
LEEVGRKRIPNAVLIPSGFAETGEQALQDEIVAVAERHGVRLLGPNIYGYYSTWQDLC
ATFCTPYDVKGGVALTSQSGGIGMAILGFARTTKTGVSAIVGLGNKSDLDEDDLLTWF
GEDPHTECIAMHLEDLKDGRAFVEAARATVPRKPVVVLKAGRTAAGAKAAGSHTGALA
GDDAVYDDILRQAGVIRAPGLNEMLEYARALPVLPAPEGDNVVIITGAGGSGVLLSDA
VTDNGLSLMEIPPDLDRAFRAFIPPFGAAGNPVDITGGEPPSTYEATIRLGLEDPRIH
ALVLGYWHTIVTPPMVFAELTARVVAEFRERGVEKPVVASLAGDVEVEEACQYLFERG
VVAYPYTTEKPVAVLGAKYRWARAAGLLGGGS"

misc_feature complement(3340. .3387)
/gene="SCO6582"
/note="PS00216 Sugar transport proteins signature 1"

gene complement(4382. .5614)
/gene="SCO6583"

CDS complement(4382. .5614)
/gene="SCO6583"
/note="SC8A6.04c, probable transferase, len: 410 aa; highly similar to TR:O06644 (EMBL:U82167) formyl-CoA transferase from Oxalobacter formigenes (428 aa), fasta scores; opt: 954 z-score: 1732.0 E(): 0, 50.7% identity in 428 aa overlap. Also similar to many dehydratases e.g. CAIB_ECOLI L-carnitine dehydratase (EC 4.2.1.89) (405 aa), fasta scores; opt: 285 z-score: 317.7 E(): 1.9e-10, 26.2% identity in 427 aa overlap"
/codon_start=1
/transl_table=11
/product="putative transferase"
/protein_id="CAA19776.1"
/db_xref="GI:3288604"
/db_xref="SPTREMBL:O87838"
/translation="MTAKALEGIRVLDMTHVQSGPSATQLLAWLGADVVKLEAPHGDI
TRGQLRDLPDVDSLYFTMLNCNKRSITLNTKSERGKEILTELIRRSDVMVENFGPGAV
DRMGFTWDRVKEINPRIVYASIKGFGEGPYTAFKAYEVVAQAMGGSMSTTGFEDGPPL
ATGAQIGDSGTGVHVVAGILAALYQREHTGRGQRVNVAMQHAVLNLCRVKLRDQQRLS
HGPLAEYPNEDFGDEVPRSGNASGGGQPGWAVKCAPGGFNDYVYVIVQPVGWQPLSEL
IGRPELAEDPEWATPRARLPKLNKMFQLIEEWSSTLPKWEVLERLNAHNIPCGPILST
KEIIEDDSLVANEMVVTVPHPERGEFVTVGSPLKLSDSPVEVTSSPLLGEHNEEVYVG
ELGLGDEELRLLKSSGVI"

RBS complement(5623. .5627)
/note="possible RBS upstream of SC8A6.04c"

gene complement(5626. .7308)
/gene="SCO6584"

CDS complement(5626. .7308)
/gene="SCO6584"
/note="SC8A6.05c, probable TPP-requiring enzyme, len: 560 a a; similar to many e.g. ILVB_MYCTU probable acetolactate synthase (EC 4.1.3.18) (547 aa), fasta scores; opt: 733 z-score: 1305.6 E(): 0, 36.4% identity in 557 aa overlap. Contains PS00187 Thiamine pyrophosphate enzymes signature and Pfam match to entry PF00205 TPP_enzymes, Thiamine pyrophosphate enzymes, score 343.90, E-value 5e-106"
/codon_start=1
/transl_table=11

```
                     /product="putative TPP-requiring enzyme"
                     /protein_id="CAA19777.1"
                     /db_xref="GI:3288605"
                     /db_xref="SPTREMBL:O87839"
                     /translation="MPDDTQDVISGGHLVAKALKAEGVDRIYTLCGGHIIDIYDGCVD
                     EGIEVVDVRHEQVAAHAADGYARITGKPGCAVVTAGPGTTDAVTGVANAFRAESPMLL
                     IGGQGAHTQHKMGSLQDLPHVDMMTPITKFAATVPDTARAADMVSMAFRECYHGAPGP
                     SFLEIPRDVLDAKVPAEKARVPRAGGYRASTRSAGDPEAVERLADLLVHAEKPAVLLG
                     SQVWTTRATEAAVELVRTLNVPAYMNGAGRGTLPPGDPHHFQLSRRYAFSNADVIVIV
                     GTPFDFRMGYGKRLSPDATVVQIDLDYRTVGKNRDIDLGIVGDAGLVLKSVTEAASGR
                     INGGAAKRKEWLDELRAAERTALDKRLPQLRSDASPIHPYRLVSEINDFLTEDSIYIG
                     DGGDIVTFSGQVVQPKSPGHWMDPGPLGTLGVGVPFVLAAKQARPDKEVVALFGDGAF
                     SLTGWDFETLVRYDLPFVGIVGNNSSMNQIRYGQAVKYGKERERVGNTLGDVPYDEFA
                     RMLGGHGEEVRDPADIGPALRRARESGKPSLINVWVDPDAYAPGTMNQTMYK"
     misc_feature    complement(5683..7254)
                     /gene="SCO6584"
                     /note="Pfam match to entry PF00205 TPP_enzymes, Thiamine
                     py rophosphate enzymes, score 343.90, E-value 5e-106"
     misc_feature    complement(5962..6021)
                     /gene="SCO6584"
                     /note="PS00187 Thiamine pyrophosphate enzymes signature"
     RBS             complement(7316..7319)

  Query Match              7.9%;  Score 197.4;  DB 1;  Length 37445;
  Best Local Similarity   47.1%;  Pred. No. 3.1e-15;
  Matches 1152;  Conservative    0;  Mismatches 1176;  Indels  119;  Gaps   13;

Qy         34 AAGCTCACCCTGGCCGAGAAGGTCGATCTGCTGGCTGGTATCGACTTCTGGCACACAAAG 93
Db      27611 AAGCTGACACCACGCGCCCGCGCACTGCTGCTGAACGGCGCCACGACCTGGCGCACGAGG 27670

Qy         94 GCTCTCCCCAAGCATGGAGTCCCCTCTCTCCGCTTTACAGATGGCCCCAACGGCGTAAGA 153
Db      27671 GCGGAACCAGCGGTGGAGCTGAGGGAGTTGGTGATGTCGGACGGTCCGGCGGGCGTACGG 27730

Qy        154 GGGACCAAGTTCTTCAATGGCGTCCCTGCGGCCTGCTTCCCTTGCGGCACGTCGCTCGGT 213
Db      27731 GGCGAGGCCTGGGACGAGCGGAGCACCTCTCTCCTGCTGCCCTCCGCCTCGGCGCTCGCC 27790

Qy        214 TCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAGAGGCCATC 273
Db      27791 GCCACCTGGGACGAGGCGCTGGTCGAAGACCTCGGTGGCCTGCTCGCCGCCGAGGCCCGG 27850

Qy        274 GCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCCCTCTCGGT 333
Db      27851 CGCAAGGGCGTGGACGTCCTCCTCGCCCCGACCCTCAACCTGCACCGCAGCCCGCTGGGC 27910

Qy        334 GGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAGCTGCGGCT 393
Db      27911 GGCCGGCACTTCGAGTGCCTCTCCGAGGACCCCGAGCTGACCGGCCGGATCGGCGCCGCG 27970

Qy        394 CTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTTTGTGCAAT 453
Db      27971 CTGGTCCGCGGGATCCAGGCGCACGGCGTGGCCGCCACCGCCAAGCACTACGTGGCCAAC 28030

Qy        454 GATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTCTCCGTGAA 513
Db      28031 GACTCCGAGACCGACCGCCTCACCGTCGACGTGCGGGTGGGCGAACGGGCGCTGCGAGAG 28090

Qy        514 ATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGTTCATGACG 573
Db      28091 GTCTACCTCGCCCCCTTCGAGGCGGCGGTG---GCCGCCGGGGTCCGGCTCGTCATGGCG 28147

Qy        574 GCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTGATGGGATG 633
Db      28148 GGGTACAACGCGGTCAACGGCACCACGATGACCGCGAACGC---CCTCCTCACCGACCCG 28204

Qy        634 CTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCACATACAGT 693
Db      28205 CTGAAGAGCGAGTGGGGCTTCGACGGCGTCGTCGTGTCCGACTGGGGCGCGGTGCGCGGC 28264

Qy        694 ACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGGACCTCCACGCTTCCGA 753
Db      28265 ACGACCGGCACCGCCCGCGCCGGTCTCGACCTCGCCATGCCGGGTCCCGACGGCCCCTGG 28324

Qy        754 GGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCTTTATCCACGTCATTGACCAG 813
Db      28325 GGCGAGGCGCTGGCCCGCGCGGTGGCCGAGGGCGCGGTGCCCGAACCGGCCGTCGACGAC 28384

Qy        814 AGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCTGCCTCCGGAGTGACGGAGAAC 873
Db      28385 AAGGCACGGCGCCTGCTGCGACTCGCGGCGTGGCTGGGCGCGCTGGGCGGGCGCGACGTG 28444

Qy        874 GGCCCCGAGACGACTGTCAACAACACCCCCGAAACGGCAGCTCTCCTCCGGAAGGTTGGC 933
Db      28445 TCCCGGTCGCCGGTCCCGGGCAGGCCGGCCGACTCGCCGGGTGCGGAGGGTGCGGACGGT 28504

Qy        934 AACGAGGGCATCGTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCAAGAAGAAG 993
Db      28505 GGGGCGGGCGCTGGCCCGTCGTCCGGTGCGGAGGGCCTCCCGGGCCGGGGCCCGGCGCAC 28564

Qy        994 AAGACGCTGATTGTCGGCCCCAACGCCAAGCAGGCCACATACCA--------------- 1037
Db      28565 GGTGCGAAGCCGTCCGGGCCCCGACCGCGGCGTGCCGGGGACGGGCGGGCGCTGGCCCGT 28624

Qy       1038 -----CGGCGGAGGCTCTGCCGCACTCAGGGCCTACTACGCAGTC-------------- 1077
Db      28625 CGTGCCGTCGCCGCCGGGGCCGTGCTGCTGGCCAACAAGGACGTCCTGCCACTCGACCCC 28684

Qy       1078 --ACTCCCTTTGACGGCCTCAGCAAGCAGCTCG-AGACGCCGCCATCGTACACCGTCGGC 1134
Db      28685 GAGCACCTCGGGACGGTCGCCGTGATCGGCGCGCACGCCGCGCGGACCCGTACCCAGGGC 28744

Qy       1135 GCCTACACCACCGTTCCTCCCATTCTAGGCGAGCAGTGCCTCACGCCCGACGGCGCTCCG 1194
Db      28745 GGCGGCAGCGCGGGCGTCTTCCCGCGGGGCGAGGTGTCCGTCCTCGACGGCATCCGGGCC 28804

Qy       1195 GGCATGCGCTGGAGGGTCTTCAACGAGCCC--CCTGGTACCCCTAACCGCCAGCACATTG 1252
Db      28805 GAACTGCGCGGCCGGGCCCGCGTCGTGCACGTCCCGGGCCCCCGGCCGGACGGCCCCGCG 28864

Qy       1253 ACGAGCTCTTCTTCACCAAGACGGACATGCACCTGGTGGACTACTACCACCCCAAGGCGG 1312
Db      28865 CCCCCACTGGACCCGGACACATGCACCGACCCGCGCTCGGGGCTGCCCGGCGTCCTGCTG 28924

Qy       1313 CAGACACGTGGTACG--CCGACATGGAGGGCACGTACACCGCCGACGAGGACTGCACCTA 1370
Db      28925 CGGATGCTCGACGCGGACGGCCGCGAGCTGTACGCCGAACGGCGCCGCGGCGGGCGCCTG 28984

Qy       1371 CGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGGCGTACGTAGACGACCAGCTCGTCGT 1430
Db      28985 CTGGAGC-CCCGCCTGGTGCCGGGCGCGCACACCGTCGAGATCCGCGCCCGGCTGTGTCC 29043

Qy       1431 CGACAACGCCACCAAGCAGGTCCCCGGCGATGCCTTCTTCGGCTCCGCCACCCGCGAGGA 1490
Db      29044 CCGCACCGGCGGCTCCTGGTCCCTGGGCGTGGCCGGGTTCGGCCGGATGAGCCTGACGAC 29103

Qy       1491 GACGGGCCGCATCAATCTCGTCAAGGGCAACACG------------------------- 1524
Db      29104 GGACGGACGCACCCTGCTGGAGGGGGACTTCCCGCCGTCCACCGACGATCCGGCGGTGAT 29163

Qy       1525 ----------------TACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACCCT 1568
Db      29164 GCACGTCAACCCGCCCGCCCAGTACGCCACCGCCGACCTCACCGCCGGCCGGGACACCCT 29223

Qy       1569 CAAGGGCGACA-----------CCATCGTCCCCGGCCACGGCTCCCTCCGCGTCGGCGG 1616
Db      29224 GCTGGTGGCCCGGCGCGAGCTGGCACCCGGCACCGGCCGGGCGACCGTCCTCGTCGCGGC 29283

Qy       1617 CTGCAAGGTCATTGACGACCAGGCCGAAATCGAAAAGTCCGTCGCCCTCGCCAAGGAGCA 1676
Db      29284 CCCGCCCGCCCCGGACGTGACCGCGTCGCTCGCCGAGGCCGTCCGCGCGGCCGGTGCGGC 29343

Qy       1677 CGACCAGGTCATCATCTGCGCGGGCCTTAACGCCGACTGGGAGACCGAGGGCGCCGACCG 1736
Db      29344 GGACGCCGCGGTCGTGGTCGTCGGCACCACCGAGCACGGGGAGTCGGAGGGCTACGACCG 29403
```

```
Qy   1737 CGCGAGCATGAAGCTCCCCGGCGTGCTGGACCAGCTCATTGCCGACGTGGCCGCCGCGAA 1796
Db  29404 TACGGACCTGGCGCTCGGCGCCACCCAGGACGCGCTGGTCCGCGCCGTCGCGGCCGCCAA 29463

Qy   1797 CCCAAACACCGTCGTCGTCATGCAGACGGGCACCCCCGAGGAGATGCCCTGGCTCGACGC 1856
Db  29464 CCCGCGCACCGTCGCCGTCGTCAACAGCGGCGGCCCGGTGGAACTGCCGTGGCGGGAGCA 29523

Qy   1857 CACGCCCGCCGTCATCCAGGCCTGGTACGGCGGCAACGAGACGGGCAACTCCATTGCCGA 1916
Db  29524 GGCGGGTGCGGTGCTGCTGGCCTGGTTTCCCGGACAGGAGGGCGGCGGTGGACTGGCCGA 29583

Qy   1917 CGTCGTCTTTGGCGACTACAACCCCTCGGGCAAGCTGTCCCTCAGCTTCCCCAAGCGCCT 1976
Db  29584 CGTGCTCTTCGGGCACGCCGAGCCGGGCGGACGACTGCCCACCACCTGGCC---GGCCGT 29640

Qy   1977 GCAGGACAACCCCGCGTTTCTCAACTTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGA 2036
Db  29641 CCTCGCCGACGCCCCGGTCACCCGCACCCGCCCCGACGGCGGCCGCCTCGACTACGACGA 29700

Qy   2037 GGACGTCTACGTCGGGTACAGGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTT 2096
Db  29701 GGGACTGCACCTCGGTCACCGGGGCTGGCTGCGCCATCACCGCACGCCCGCCTACTGGTT 29760

Qy   2097 TGGCCACGGCCTGTCCTACACCACTT---------TTGCCTTTTCCAATCTCTCCGTGTC 2147
Db  29761 CGGACACGGGCTCGGCTACACGACGTGGCGGTACGAGGAGCTGACCGTCCCGCCGGTGAC 29820

Qy   2148 TCACAAGGACGGCAAGCTGAGCGTGTCCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGG 2207
Db  29821 CCGGGCGGGCGACGGCCTCACCGTGCGCGTGCGGGTGCGCAACACCGGTGCGCGAGCGGG 29880

Qy   2208 CGCACAGGTGGCCCAGCTCTACGTCAAGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGT 2267
Db  29881 CCGGGAGGTCGTCCAGGTGTACCT---GGCCCGGCCCGCGTCGGCCCTCGACCGTCCCGC 29937

Qy   2268 CAAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTGCAGCCCGGCGAGACGAAGGCGGTGAC 2327
Db  29938 GCGTTGGCTCGCCGGGTACACGGCGGTGCGGGCGCGCCCGGGGGAGACGGTGACGGCGAC 29997

Qy   2328 AATCGAGGAGCAGGAGAAGTACGTCGCTGCGTATTTTGATGAGGAGC 2374
Db  29998 GGTGCGCGTCCCGGCGCGGGCCCTGCGCCACTGGTCGGTGGCGGAGC 30044
```

```
RESULT 13
AR173232
LOCUS       AR173232                3241 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 11 from patent US 6303767.
ACCESSION   AR173232
VERSION     AR173232.1  GI:17912723
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3241)
  AUTHORS   Betlach,M.C. and McDaniel,R.
  TITLE     Nucleic acids encoding narbonolide polyketide synthase enzymes from
            streptomyces narbonensis
  JOURNAL   Patent: US 6303767-A 11 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .3241
                     /organism="unknown"
BASE COUNT      520 a   1174 c   1125 g    422 t
ORIGIN

  Query Match             7.4%;  Score 184.4;  DB 6;  Length 3241;
  Best Local Similarity   48.2%;  Pred. No. 2.1e-13;
  Matches 1028;  Conservative    0;  Mismatches  956;  Indels  150;  Gaps   12;

Qy    206 CGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGGCAAAG 265
```

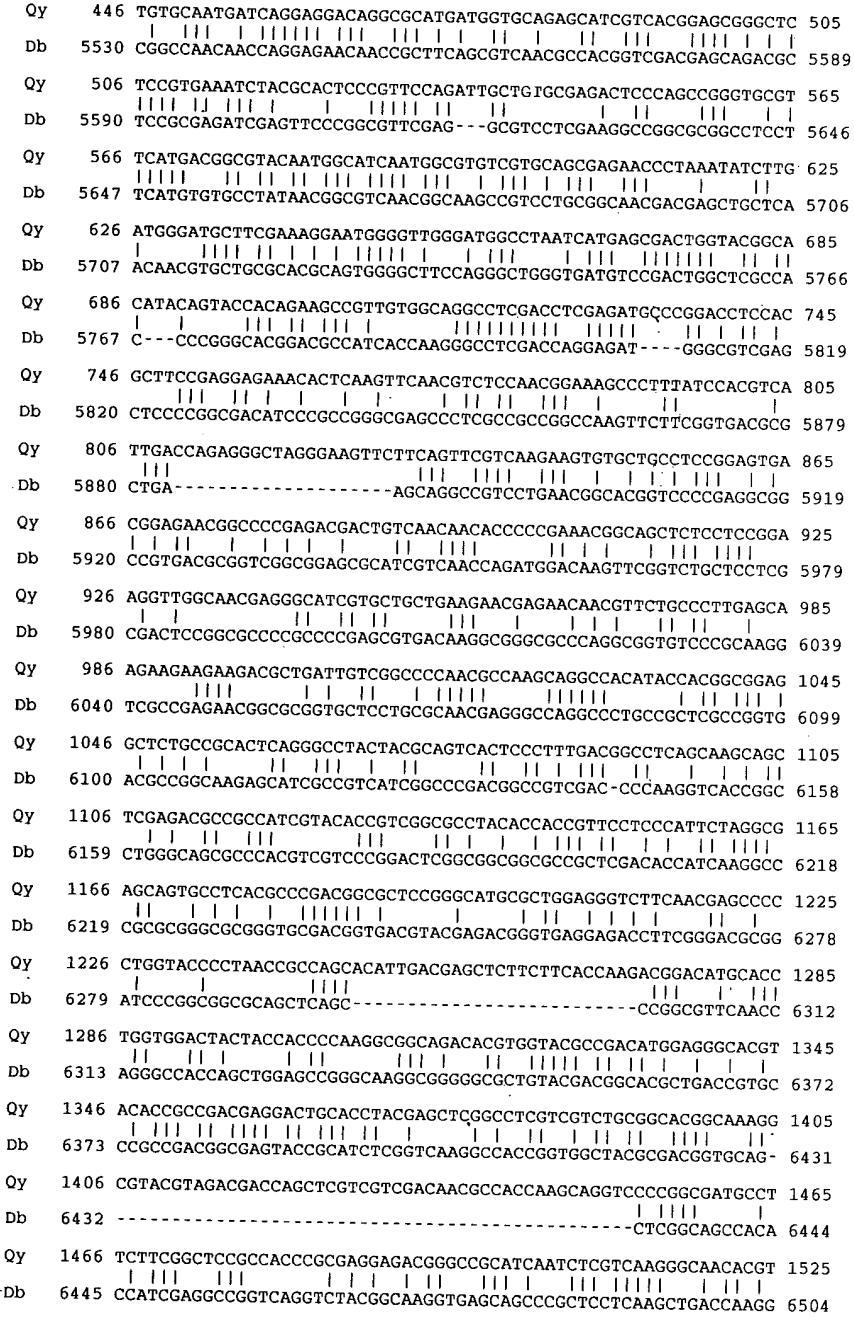
```
Db   1151 CCCTGGCCAGCACCTTCGACGACTCCATGGCCGACAGCTACGGCAGGGTCATGGGCCGCG 1210

Qy    266 AGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCC 325
Db   1211 ACGGACGCGCGCTGGGCCAGGACATGGTTCTGGGCCCGATGATGAACAACATCCGGGTGC 1270

Qy    326 CTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAG 385
Db   1271 CACACGGCGGCCGGAACTACGAGACCTTCAGCGAGGACCCCCTGGTCTCCTCGCGCACCG 1330

Qy    386 CTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTT 445
Db   1331 CGGTCGCCCAGATCAAGGGCATCCAGGGTGCGGGTCTGATGACCACGGCCAAGCACTTCG 1390

Qy    446 TGTGCAATGATCAGGAGGACAGGCGCATGATGGTGCAGAGCATCGTCACGGAGCGGGCTC 505
Db   1391 CGGCCAACAACCAGGAGAACAACCGCTTCAGCGTCAACGCCACGGTCGACGAGCAGACGC 1450

Qy    506 TCCGTGAAATCTACGCACTCCCGTTCCAGATTGCTGTGCGAGACTCCCAGCCGGGTGCGT 565
Db   1451 TCCGCGAGATCGAGTTCCCGGCGTTCGAG---GCGTCCTCGAAGGCCGGCGCGGCCTCCT 1507

Qy    566 TCATGACGGCGTACAATGGCATCAATGGCGTGTCGTGCAGCGAGAACCCTAAATATCTTG 625
Db   1508 TCATGTGTGCCTATAACGGCGTCAACGGCAAGCCGTCCTGCGGCAACGACGAGCTGCTCA 1567

Qy    626 ATGGGATGCTTCGAAAGGAATGGGGTTGGGATGGCCTAATCATGAGCGACTGGTACGGCA 685
Db   1568 ACAACGTGCTGCGCACGCAGTGGGGCTTCCAGGGCTGGGTGATGTCCGACTGGCTCGCCA 1627

Qy    686 CATACAGTACCACAGAAGCCGTTGTGGCAGGCCTCGACCTCGAGATGCCCGGACCTCCAC 745
Db   1628 C---CCCGGGCACGGACGCCATCACCAAGGGCCTCGACCAGGAGAT----GGGCGTCGAG 1680

Qy    746 GCTTCCGAGGAGAAACACTCAAGTTCAACGTCTCCAACGGAAAGCCCTTTATCCACGTCA 805
Db   1681 CTCCCCGGCGACATCCCGCCGGGCGAGCCCTCGCCGCCGGCCAAGTTCTTCGGTGACGCG 1740

Qy    806 TTGACCAGAGGGCTAGGGAAGTTCTTCAGTTCGTCAAGAAGTGTGCTGCCTCCGGAGTGA 865
Db   1741 CTGA--------------------AGCAGGCCGTCCTGAACGGCACGGTCCCCGAGGCGG 1780

Qy    866 CGGAGAACGGCCCCGAGACGACTGTCAACAACACCCCCGAAACGGCAGCTCTCCTCCGGA 925
Db   1781 CCGTGACGCGGTCGGCGGAGCGCATCGTCAACCAGATGGACAAGTTCGGTCTGCTCCTCG 1840

Qy    926 AGGTTGGCAACGAGGGCATCGTGCTGCTGAAGAACGAGAACAACGTTCTGCCCTTGAGCA 985
Db   1841 CGACTCCGGCGCCCCGCCCCGAGCGTGACAAGGCGGGCGCCCAGGCGGTGTCCCGCAAGG 1900

Qy    986 AGAAGAAGAAGACGCTGATTGTCGGCCCCAACGCCAAGCAGGCCACATACCACGGCGGAG 1045
Db   1901 TCGCCGAGAACGGCGCGGTGCTCCTGCGCAACGAGGGCCAGGCCCTGCCGCTCGCCGGTG 1960

Qy   1046 GCTCTGCCGCACTCAGGGCCTACTACGCAGTCACTCCCTTTGACGGCCTCAGCAAGCAGC 1105
Db   1961 ACGCCGGCAAGAGCATCGCCGTCATCGGCCCGACGGCCGTCGAC-CCCAAGGTCACCGGC 2019

Qy   1106 TCGAGACGCCGCCATCGTACACCGTCGGCGCCTACACCACCGTTCCTCCCATTCTAGGCG 1165
Db   2020 CTGGGCAGCGCCCACGTCGTCCCGGACTCGGCGGCGGCGCCGCTCGACACCATCAAGGCC 2079

Qy   1166 AGCAGTGCCTCACGCCCGACGGCGCTCCGGGCATGCGCTGGAGGGTCTTCAACGAGCCCC 1225
Db   2080 CGCGCGGGCGCGGGTGCGACGGTGACGTACGAGACGGGTGAGGAGACCTTCGGGACGCGG 2139

Qy   1226 CTGGTACCCCTAACCGCCAGCACATTGACGAGCTCTTCTTCACCAAGACGGACATGCACC 1285
Db   2140 ATCCCGGCGGCGCAGCTCAGC------------------------CCGGCGTTCAACC 2173

Qy   1286 TGGTGGACTACTACCACCCCAAGGCGGCAGACACGTGGTACGCCGACATGGAGGGCACGT 1345
Db   2174 AGGGCCACCAGCTGGAGCCGGGCAAGGCGGGGGCGCTGTACGACGGCACGCTGACCGTGC 2233
```

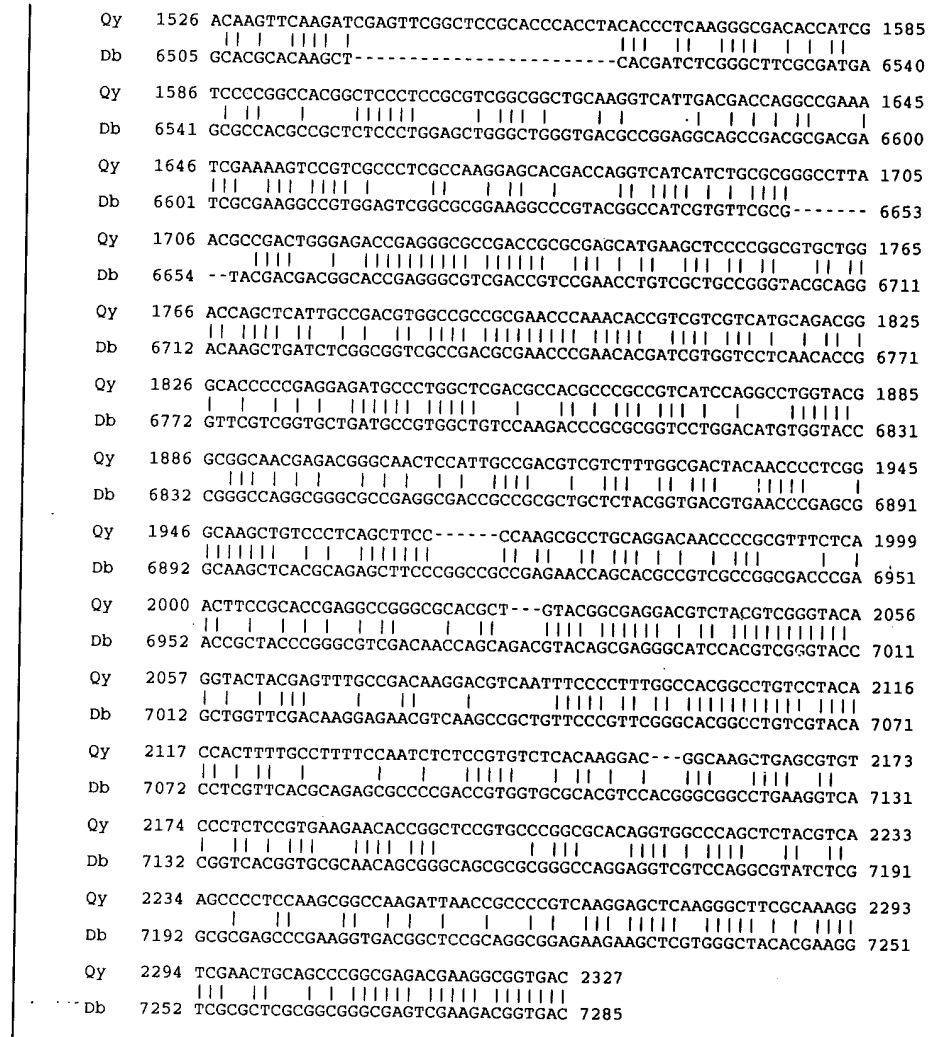
```
Qy   1346 ACACCGCCGACGAGGACTGCACCTACGAGCTCGGCCTCGTCGTCTGCGGCACGGCAAAGG 1405
          | ||| || |||| || ||| ||  |      | |  ||  | || ||  ||||   ||
Db   2234 CCGCCGACGGCGAGTACCGCATCTCGGTCAAGGCCACCGGTGGCTACGCGACGGTGCAG- 2292

Qy   1406 CGTACGTAGACGACCAGCTCGTCGTCGACAACGCCACCAAGCAGGTCCCCGGCGATGCCT 1465
                                                          | ||||  .  |
Db   2293 ----------------------------------------------CTCGGCAGCCACA 2305

Qy   1466 TCTTCGGCTCCGCCACCCGCGAGGAGACGGGCCGCATCAATCTCGTCAAGGGCAACACGT 1525
          | |||  |||        | | |  | ||   ||| |   ||| ||||   | || | .
Db   2306 CCATCGAGGCCGGTCAGGTCTACGGCAAGGTGAGCAGCCCGCTCCTCAAGCTGACCAAGG 2365

Qy   1526 ACAAGTTCAAGATCGAGTTCGGCTCCGCACCCACCTACACCCTCAAGGGCGACACCATCG 1585
          || |  |||| |                    |||  ||  ||||  | | ||
Db   2366 GCACGCACAAGCT----------------------CACGATCTCGGGCTTCGCGATGA 2401

Qy   1586 TCCCCGGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAA 1645
          | ||   |   ||||||     | ||| |    | | |       |  | | | ||   |
Db   2402 GCGCCACGCCGCTCTCCCTGGAGCTGGGCTGGGTGACGCCGGAGGCAGCCGACGCGACGA 2461

Qy   1646 TCGAAAAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTA 1705
          |||  ||| |||| |     ||   | ||  |     || |||| | | ||||
Db   2462 TCGCGAAGGCCGTGGAGTCGGCGCGGAAGGCCCGTACGGCCATCGTGTTCGCG------- 2514

Qy   1706 ACGCCGACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGG 1765
             ||||   |  |||||||||| ||||||  ||| | ||  ||| || ||   || ||
Db   2515 --TACGACGACGGCACCGAGGGCGTCGACCGTCCGAACCTGTCGCTGCCGGGTACGCAGG 2572

Qy   1766 ACCAGCTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGG 1825
          || |||| ||  | |  || |||| |||||||||| |||||  |||| ||| |  | || |
Db   2573 ACAAGCTGATCTCGGCGGTCGCCGACGCGAACCCGAACACGATCGTGGTCCTCAACACCG 2632

Qy   1826 GCACCCCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACG 1885
          |  |  | | |  |||||| |||||   |   || | ||| ||| |  |    ||||||
Db   2633 GTTCGTCGGTGCTGATGCCGTGGCTGTCCAAGACCCGCGCGGTCCTGGACATGTGGTACC 2692

Qy   1886 GCGGCAACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGG 1945
             ||| | | |   | | |  | |  ||||   |  |||  || |||   |||||    |
Db   2693 CGGGCCAGGCGGGCGCCGAGGCGACCGCCGCGCTGCTCTACGGTGACGTGAACCCGAGCG 2752

Qy   1946 GCAAGCTGTCCCTCAGCTTCC------CCAAGCGCCTGCAGGACAACCCCGCGTTTCTCA 1999
          |||||||  | |  |||||||      || ||  || ||| | |  | |||     |  |
Db   2753 GCAAGCTCACGCAGAGCTTCCCGGCCGCCGAGAACCAGCACGCCGTCGCCGGCGACCCGA 2812

Qy   2000 ACTTCCGCACCGAGGCCGGGCGCACGCT---GTACGGCGAGGACGTCTACGTCGGGTACA 2056
          ||  |  | | |  | ||    |  ||    |||| |||||| | || ||||||||||
Db   2813 ACCGCTACCCGGGCGTCGACAACCAGCAGACGTACAGCGAGGGCATCCACGTCGGGTACC 2872

Qy   2057 GGTACTACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACA 2116
          | |  | |||    |  ||    |       ||||| || || ||||||||||| ||||
Db   2873 GCTGGTTCGACAAGGAGAACGTCAAGCCGCTGTTCCCGTTCGGGCACGGCCTGTCGTACA 2932

Qy   2117 CCACTTTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGAC---GGCAAGCTGAGCGTGT 2173
          || | ||  |      |   |  |||||   | || |  |   |||  ||||  ||
Db   2933 CCTCGTTCACGCAGAGCGCCCCGACCGTGGTGCGCACGTCCACGGGCGGCCTGAAGGTCA 2992

Qy   2174 CCCTCTCCGTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCA 2233
          |  || | |||   |||| |||        | |||    |||| | ||||   ||  ||
Db   2993 CGGTCACGGTGCGCAACAGCGGGCAGCGCGCGGGCCAGGAGGTCGTCCAGGCGTATCTCG 3052

Qy   2234 AGCCCCTCCAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGG 2293
             |  ||   ||  | |  |   |  | |  ||| ||||| |||||  | | ||||
Db   3053 GCGCGAGCCCGAAGGTGACGGCTCCGCAGGCGGAGAAGAAGCTCGTGGGCTACACGAAGG 3112

Qy   2294 TCGAACTGCAGCCCGGCGAGACGAAGGCGGTGAC 2327
          |||  ||   | | |||||| ||||| |||||||
Db   3113 TCGCGCTCGCGGCGGGCGAGTCGAAGACGGTGAC 3146


RESULT 14
AF521878
LOCUS       AF521878               17665 bp    DNA     linear   BCT 29-JUL-2002
```

```
DEFINITION  Streptomyces narbonensis desosamine biosynthetic gene cluster,
            partial sequence.
ACCESSION   AF521878
VERSION     AF521878.1  GI:22001230
KEYWORDS    .
SOURCE      Streptomyces narbonensis.
  ORGANISM  Streptomyces narbonensis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 17665)
  AUTHORS   Bate,N. and Cundliffe,E.
  TITLE     The desosamine biosynthetic cluster of Streptomyces narbonensis,
            producer of narbomycin
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 17665)
  AUTHORS   Bate,N. and Cundliffe,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2002) Biochemistry, Leicester University,
            University Road, Leicester LE1 7RH, UK
FEATURES             Location/Qualifiers
     source          1. .17665
                     /organism="Streptomyces narbonensis"
                     /db_xref="taxon:67333"
     gene            <1. .528
                     /gene="nbmA"
     CDS             <1. .528
                     /gene="nbmA"
                     /note="terminal thioesterase I (TEI) domain of type I
                     polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NbmA"
                     /protein_id="AAM88350.1"
                     /db_xref="GI:22001231"
                     /translation="AVLRAAGDAPVVLLGHSGGALLAHELAFRLERAHGTAPAGIVLV
                     DPYPPGHQEPIEVWSRQLGEGLFAGELEPMSDARLLAMGRYARFLAGPRPGRSSAPVL
                     LVRASERLGDWPEERGDWRAHWDLPHSVADVPGDHFTMMRDHAPAVAVAVLAWLDGIE
                     RDTAAGRRTAQGADQ"
     gene            525. .1370
                     /gene="nbmB"
     CDS             525. .1370
                     /gene="nbmB"
                     /note="thioesterase II (TEII)"
                     /codon_start=1
                     /transl_table=11
                     /product="NbmB"
                     /protein_id="AAM88351.1"
                     /db_xref="GI:22001232"
                     /translation="MTDRVLNVDSSLWIRRFHPSPNSAVRLVCLPHAGGSASYFFRFS
                     EELHPSVEALSVQYPGRQDRRAEPCLESVEELAEHVVAATEAWWREGRLAFFGHSLGA
                     SVAYEAARILEQRHGVRPEGLYVSGRRAPSLAPDRLVHCLDDRAFLAEIRQLSGTDER
                     FLQDDELMRLVLPALRSDYKAAETYEHRPSAKLACPVMALAGDRDPKAPLNEVAEWRR
                     HTSGPFCLRAYSGGHFYLNDQWHGICNDISDHLLVTRGALDSRVVQPPTRLIEGAAKR
                     WQNPR"
     gene            1367. .2713
                     /gene="nbmC"
     CDS             1367. .2713
                     /gene="nbmC"
                     /note="putative NDP-deoxyhexose 3,4-isomerase"
                     /codon_start=1
                     /transl_table=11
                     /product="NbmC"
                     /protein_id="AAM88352.1"
                     /db_xref="GI:22001233"
                     /translation="MTDDPARALTQPPLGRTVSAVADPELGAHLLETRGIHWIHAANG
                     DPYAAVLRGQADDPYPAYERVRARGVLSFSPTGSWVTADHAVAASVLCSTEFGVPGAD
                     GVPVPQQVLSYGEGCPLEHEQAPRAAAEVPGAWAGQRRAVVEEVHRQILEGLAPDPSA
                     SYAFELVGGFVRPAVTGAAAAVLGVPAGRRGEFAELVERLRPLSDSLLAPQSLRTVRA
                     ADGALAALTALLAEAGVSAATDGSDPSSGPSAPSSDPSASSSGPTAPSSGPTAPSSGV
                     AGRPDGALLSALGVTAAVQLTGNAVLALLAHPEQWRELCARPGLAAAAVEETLRYDPP
                     VQLDARVARVETELAGRRLPAGAHVVVLTAGTGRDPGVFTDPERFDLARPDATAHLAL
                     HPAGPYGPVASLVRLQAEVALRTLAGRFPGLRKAGPVLRPRRAPVGRGPLSVPVGS"
     gene            2782. .4062
```

```
                /gene="nbmD"
CDS             2782. .4062
                /gene="nbmD"
                /note="desosaminyl transferase"
                /codon_start=1
                /transl_table=11
                /product="NbmD"
                /protein_id="AAM88353.1"
                /db_xref="GI:22001234"
                /translation="MRVLLTSFAHHTHYYGLVPLAWALLAAGHEVRVASQPALTDTIT
                GSGLAAVPVGTDHLIHEYRVRMAGEPRPNHPAIAFDEARSEPLDWDHALGIDAILAPY
                FYLLANNDSMIDDLVGFARSWQPDLVLWEPTTYAGAVAAQVTGAAHARVLWGPDVMGS
                ARRKFVALRDRQPPEHREDPTAEWLTWTLDRFGASFEEELLTGQFTVDPTPPSLRLDT
                GLPTVGVQYVPYNGTSVVPDWLAEPPARPRVCLTLGVSAREVLGGDGVSQADILEALA
                DLDIELVATLDAGQRAEIRSYPKHTRFTDFVPMHALLPSCSAIVHHGGAGTYATAVIN
                AVPQVMLAELWDAPVKARAVAEQGAGFFLPPAGLTPQAVRDAVVRILDDPSVLAAAHR
                LREETHRDPTPAGIVPELERLAAQHRRPPADARR"
gene            4151. .4864
                /gene="nbmE"
CDS             4151. .4864
                /gene="nbmE"
                /note="N,N-methyltransferase"
                /codon_start=1
                /transl_table=11
                /product="NbmE"
                /protein_id="AAM88354.1"
                /db_xref="GI:22001235"
                /translation="MYEVDHADVYDLFYLGRGKDYAAEASDIADLVRTRTPEASSLLD
                VACGTGTHLEHFTKEFGDTAGLELSEDMLTHARKRLPDATLHQGDMRDFRLGRRFSAV
                VSMFSSVGYLRTTAELDAAVASFAAHLEPGGVVVVEPWWFPETFADGWVSADVVRRDG
                RTVARVSHSVRDGDATRMEVHFTVADPGRGVRHFSDVHLITLFHRAEYEAAFTAAGLR
                VEYLEGGPSGRGLFVGVPA"
gene            4908. .7424
                /gene="nbmF"
CDS             4908. .7424
                /gene="nbmF"
                /note="beta glucosidase"
                /codon_start=1
                /transl_table=11
                /product="NbmF"
                /protein_id="AAM88355.1"
                /db_xref="GI:22001236"
                /translation="MHQAQREKSTVTGKTRIPRVRRSRTTPRAFTLAVVGTLLAGTTV
                AAAAPGAAGTGHVQYTSKAAELVAQMTLDEKISFVHWALDPDRQNVGYLPGVPRLGIP
                ELRAADGPNGIRLVGRTATALPAPVALASTFDDSMADSYGRVMGRDGRALGQDMVLGP
                MMNNIRVPHGGRNYETFSEDPLVSSRTAVAQIKGIQGAGLMTTAKHFAANNQENNRFS
                VNATVDEQTLREIEFPAFEASSKAGAASFMCAYNGVNGKPSCGNDELLNNVLRTQWGF
                QGWVMSDWLATPGTDAITKGLDQEMGVELPGDIPPGEPSPPAKFFGDALKQAVLNGTV
                PEAAVTRSAERIVNQMDKFGLLLATPAPRPERDKAGAQAVSRKVAENGAVLLRNEGQA
                LPLAGDAGKSIAVIGPTAVDPKVTGLGSAHVVPDSAAAPLDTIKARAGAGATVTYETG
                EETFGTRIPAAQLSPAFNQGHQLEPGKAGALYDGTLTVPADGEYRISVKATGGYATVQ
                LGSHTIEAGQVYGKVSSPLLKLTKGTHKLTISGFAMSATPLSLELGWVTPEAADATIA
                KAVESARKARTAIVFAYDDGTEGVDRPNLSLPGTQDKLISAVADANPNTIVVLNTGSS
                VLMPWLSKTRAVLDMWYPGQAGAEATAALLYGDVNPSGKLTQSFPAAENQHAVAGDPN
                RYPGVDNQQTYSEGIHVGYRWFDKENVKPLFPFGHGLSYTSFTQSAPTVVRTSTGGLK
                VTVTVRNSGQRAGQEVVQAYLGASPKVTAPQAEKKLVGYTKVALAAGESKTVTVNVDR
                RQLQYWDAASDSWRTGTGSRLLQTGSSSADLKNSSQVTVW"
gene            complement(7514. .8653)
                /gene="nbmG"
CDS             complement(7514. .8653)
                /gene="nbmG"
                /note="aminotransferase"
                /codon_start=1
                /transl_table=11
                /product="NbmG"
                /protein_id="AAM88356.1"
                /db_xref="GI:22001237"
                /translation="MTARTETGRVPFLDLKAAYEELRPETDAAIARVLDSGRYLLGPE
                LEAFEAEFATYCETDHAVGVNSGMDALQLALRGLGIGPGDEVIVPSHTYIASWLAVSA
                TGATPVPVEPNEDHPTLDPLLVEKAITPRTRALLPVHLYGHPADLDALRDLADRHGLH
                IVEDAAQAHGARYRGRRIGAGASVAAFSFYPGKNLGCFGDGGAVVTGDPALAERLRML
                RNYGSRQKYSHETKGTNSRLDEIQAAVLRIRLAHLDSWNGRRSTLAAEYLSGLAGLPG
                IALPVTAPGTDPVWHLFTVRTDRRDELRDHLDSRGVDTLTHYPVPVHLSPAYAGQAPP
                EGSLPRAESFARQVLSLPIGPHLERTQALRVIDAVREWAEQAGRP"
gene            complement(8650. .9684)
                /gene="nbmH"
CDS             complement(8650. .9684)
                /gene="nbmH"
                /note="NDP-glucose 4,6 dehydratase"
                /codon_start=1
                /transl_table=11
                /product="NbmH"
                /protein_id="AAM88357.1"
                /db_xref="GI:22001238"
                /translation="MRLLVTGGAGFIGSHFVRQLLAGAYPDVPAGEVIVLDSLTYAGN
                RANLAPVEADPRLRFVHGDIRDADLLARELRGVDAVVHFAAESHVDRSIAGASVFTGT
                NVQGTQTLLQCALDAGVGRVVHVSTDEVYGSIDSGSWNESSPLEPNSPYAASKAGSDL
                VARAYHRTYGLDVRITRCCNNYGPYQHPEKLIPLFVTNLLDGGTLPLYGDGANVREWV
                HTDDHCRGIALVLAGGRAGEVYHIGGGRELTNRELTGILLDSLGADWSSVRRVADRQG
                HDLRYSLDGGKIERELGYRPQIAFADGLARTIRWYRENRDWWEPLKATAPQPPATAPQ
                PPATAVGVSA"
gene            complement(9759. .10637)
                /gene="nbmI"
CDS             complement(9759. .10637)
                /gene="nbmI"
                /note="NDP-glucose synthase"
                /codon_start=1
                /transl_table=11
                /product="NbmI"
                /protein_id="AAM88358.1"
                /db_xref="GI:22001239"
                /translation="MKGIVLAGGSGTRLHPATSVISKQILPVYNKPMIYYPLSVLMLG
                GIREIQIISTPQHIELFQSLLGNGRHLGIELDYAVQKEPAGIADALLVGAGHIGDDTC
                ALILGDNIFHGPGLYSLLRDSIARLDGCVLFGYPVKDPERYGVAEVDATGRLTGLVEK
                PEKPRSHLAVTGLYLYDNDVVDIAKNIRPSPRGELEITDVNRVYLERGRAELVDLGRG
                FAWLDTGTHDSLLRAAQYVQVLEERQGVWIAGLEEIAFRMGFIDAEACHSLGEGLSRT
                EYGSYLMEIAGREGAP"
gene            complement(10840. .12297)
                /gene="nbmJ"
CDS             complement(10840. .12297)
                /gene="nbmJ"
                /note="putative NDP-deoxyhexose 3,4-reductase"
                /codon_start=1
                /transl_table=11
                /product="NbmJ"
                /protein_id="AAM88359.1"
                /db_xref="GI:22001240"
                /translation="MTAPAPVATAPSARCAHPGADLGAAVRAVGQQSLAAGGLVPSDEA
                GSTARHLVGLAVRYGNSPFTPLEEARHDLGVDRDAFRRLLALFPRVPELRAAVEGGPA
                GAYWQNTLLPLERRGVFDSALAGKPVFPYSVGLYPGPTCMFRCHFCVRVTGARYDPSA
                LDTGNAMFRSVIDEIPAGNPHALYFSGGLEPLTNPGLGTLAAHAAEHGLRPTVYTNSF
                ALTERTLERQPGLWGLHAVRTSLYGLNDEEYEQTTGKKAAFRRVRENLRRFQQLRAER
                EAPINLGFAYIVLPGRAHRLLDLVDFIADLDESGQGRTIDFVNIREDYSGREDGKLPQ
                EERAELQEALAAFEERVRERTPGLHIDYGYALNSLKAGADAELLRIKPATMRPTAHPQ
                VAVQVDLLGDVYLYREAGFPDLDGATRYIAGRVGPGTSLTEVVRDYVERGGRVAARDG
                DEYFMDGFDQVVTARLNQLERDIADGWEEARGFLR"
```

```
Query Match             7.4%;  Score 184.4;  DB 1;  Length 17665;
Best Local Similarity   48.2%;  Pred. No. 1.4e-13;
Matches 1028;  Conservative    0;  Mismatches  956;  Indels  150;  Gaps   12;

Qy    206 CGCTCGGTTCCACATTCAACCAAACTCTGCTCGAAGAGGCAGGTAAGATGATGGCGCAAAG 265
          | || |    ||| ||| || |  |  ||  |||        || | | | ||||||   |
Db   5290 CCCTGGCCAGCACCTTCGACGACTCCATGGCCGACAGCTACGGCAGGGTCATGGGCCGCG 5349

Qy    266 AGGCCATCGCTAAGAGTGCGCATGTGATCCTCGGCCCGACTATCAACATGCAACGCTCCC 325
          | |    |||   | |   | |  || | || |||||||  || ||||    ||     |
Db   5350 ACGGACGCGCGCTGGGCCAGGACATGGTTCTGGGCCCGATGATGAACAACATCCGGGTGC 5409

Qy    326 CTCTCGGTGGACGTGGCTTCGAGTCGATTGGTGAGGATCCGTTCCTGGCGGGCTTGGGAG 385
          | | ||| || ||   || |||| |  |  | ||||| ||  |  |  |          |
Db   5410 CACACGGCGGCCGGAACTACGAGACCTTCAGCGAGGACCCCCTGGTCTCCTCGCGCACCG 5469

Qy    386 CTGCGGCTCTCATCCGCGGCATTCAGAGCACTGGAGTGCAGGCTACGATCAAGCACTTTT 445
          | |  || |  |||   ||||| ||| |  | ||  ||  | | |||  |||||||||
Db   5470 CGGTCGCCCAGATCAAGGGCATCCAGGGTGCGGGTCTGATGACCACGGCCAAGCACTTCG 5529
```





RESULT 15
SEY14327/c
LOCUS SEY14327 2760 bp DNA linear BCT 11-JUN-1998
DEFINITION Saccharopolyspora erythraea partial ermE and eryBIII genes and eryB gene.
ACCESSION Y14327
VERSION Y14327.1 GI:2982610
KEYWORDS beta-glucosidase; eryBI gene; eryBIII gene.
SOURCE Saccharopolyspora erythraea.
ORGANISM Saccharopolyspora erythraea
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
REFERENCE 1 (bases 1 to 2760)
AUTHORS Gaisser,S., Bohm,G.A., Doumith,M., Raynal,M.C., Dhillon,N., Cortes,J. and Leadlay,P.F.

```
  TITLE     Analysis of eryBI, eryBIII and eryBVII from the erythromycin
            biosynthetic gene cluster in Saccharopolyspora erythraea
  JOURNAL   Mol. Gen. Genet. 258 (1-2), 78-88 (1998)
  MEDLINE   98273631
   PUBMED   9613575
REFERENCE   2  (bases 1 to 2760)
  AUTHORS   Leadlay,P.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1997) P.F. Leadlay, Department Of Biochemistry,
            University Of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK
  REMARK    Updated by [3]
REFERENCE   3  (bases 1 to 2760)
  AUTHORS   Leadlay,P.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1998) P.F. Leadlay, Department Of Biochemistry,
            University Of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK
COMMENT     On Mar 25, 1998 this sequence version replaced gi:2765397.
            Overlapping sequences: X60379, M11200.
FEATURES             Location/Qualifiers
     source          1. .2760
                     /organism="Saccharopolyspora erythraea"
                     /strain="NRRL2338"
                     /db_xref="taxon:1836"
     gene            1. .45
                     /gene="ermE"
     CDS             <1. .45
                     /gene="ermE"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA74701.1"
                     /db_xref="GI:2765398"
                     /translation="TGGRPGRRGGPGQR"
     gene            complement(100. .2526)
                     /gene="eryBI"
     CDS             complement(100. .2526)
                     /gene="eryBI"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-glucosidase"
                     /protein_id="CAA74702.1"
                     /db_xref="GI:2982611"
                     /db_xref="SPTREMBL:O70021"
                     /translation="MTGGERVKRLVIRIAPLLLVVPLLVAAVSPVRHSQRVDELIGQL
                     TLDEKLSFVYWDYNEKDPLAKLWLPGVPRLGIPQIRGTDGPAGVTIHQPAIAMPAPVA
                     LASAFDDRLAHEYGTVLGREGRAFEQDIILGPMVNNIRVPQAGRNFETFSEDPLVTAR
                     TAAAQIRGIHSQGLMTSAKHYAANTQETDRFTIDVDVDQRTLRELELPGFEAAVAAGA
                     TSVMCAYPKVNGTHACGHRQLLTEILKEQWGFKGWVMSDWTATHATEDLVAGLDQEMG
                     VEVREDGSLFRGKYLGEALKKAIREGRIPESALDASVRRILTQFERFGLLDETKPPRP
                     ERDVAGGTRIAQEVAESGAVLLRNEGGVLPLDPAAGQDIAVIGPSAQQPKVTGLGSSY
                     VEPDFANAPLDTITQRVGSGGRVGYSVGEEIKGAPIPETALQPAFVPGEVTPPPSGGV
                     IYDGRLKVDADGLYRIAARIDGGNGSLQIDGGAPIGVGDVFGPLTSVPVWLTKGEHTI
                     QMTGAAPVGGGSLDVDLTWVTPGHAQREFDAAVERARDSDVAVVFAYDDGAETADRTS
                     LSLPGTQDKLIDAVASVNPNTVVVLNTGSSVTMPWLDKTRAVLDMWYPGQAGAEATTA
                     LLFGDAEPGGRLTQTFPVSQERTPVGGDPARFPGVDGKVHYSEGIFSGYRWYDREGVD
                     PLFPFGHGLSYTTFERTDPVVERTRDGLDVTVTVRNTGQRRGSDVVQVYLGPSPQVPL
                     DQAPRQLAGYQKVELAPGETKRVRVHVAERALQHWDEAAGGWKLGGGKRAVEIGSSSR
                     DIDLRADINL"
     gene            2716. .2760
                     /gene="eryBIII"
     CDS             2716. .>2760
                     /gene="eryBIII"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA74703.1"
                     /db_xref="GI:2765400"
                     /db_xref="SPTREMBL:Q00463"
                     /translation="MIFLVGLGKCRICGN"
BASE COUNT      383 a    966 c    982 g    429 t
ORIGIN

  Query Match             7.1%;  Score 178.2;  DB 1;  Length 2760;
  Best Local Similarity   54.0%;  Pred. No. 1.2e-12;
  Matches  416;  Conservative    0;  Mismatches  343;  Indels   12;  Gaps    2;
```

```
Qy   1591 GGCCACGGCTCCCTCCGCGTCGGCGGCTGCAAGGTCATTGACGACCAGGCCGAAATCGAA 1650
          ||| ||| || |||  ||||| |     |  |||||     |  || ||| |   |||
Db    978 GGCGGCGGTTCGCTCGACGTCGACCTGACCTGGGTCACCCCGGGGCACGCCCAGCGCGAG 919

Qy   1651 AAGTCCGTCGCCCTCGCCAAGGAGCACGACCAGGTCATCATCTGCGCGGGCCTTAACGCC 1710
             ||  ||  | |    || || ||||   | | ||      |    |   | ||| |
Db    918 TTCGACGCGGCGGTGGAGCGGGCGCGCGACTCCGACGTCGCGGTGGTTTTCGCCTACGAC 859

Qy   1711 GACTGGGAGACCGAGGGCGCCGACCGCGCGAGCATGAAGCTCCCCGGCGTGCTGGACCAG 1770
          ||| | | |   |||  |||||||||| ||    |||  || ||||||    | |||| ||
Db    858 GACGGCGCG---GAGACCGCCGACCGCACGTCGCTGAGCCTGCCCGGCACCCAGGACAAG 802

Qy   1771 CTCATTGCCGACGTGGCCGCCGCGAACCCAAACACCGTCGTCGTCATGCAGACGGGCACC 1830
          || || | || ||| ||| | |  ||||| |||||||||||| ||  |  | || ||| |
Db    801 CTGATCGACGCCGTCGCCTCGGTCAACCCGAACACCGTCGTGGTGCTCAACACCGGCTCG 742

Qy   1831 CCCGAGGAGATGCCCTGGCTCGACGCCACGCCCGCCGTCATCCAGGCCTGGTACGGCGGC 1890
            | |    |||||| ||||| |||   || | ||| ||  || |    ||||||   ||
Db    741 TCGGTCACGATGCCGTGGCTGGACAAGACCCGCGCGGTGCTCGACATGTGGTACCCGGGA 682

Qy   1891 AACGAGACGGGCAACTCCATTGCCGACGTCGTCTTTGGCGACTACAACCCCTCGGGCAAG 1950
            | |    ||   |  |||   | |   | | || ||||||  | | |||   |||  |
Db    681 CAGGCCGGGGCGGAGGCCACCACGGCGCTGCTGTTCGGCGACGCCGAGCCCGGCGGCCGG 622

Qy   1951 CTGTCCCTCAGCTTCCCCAAGCGCCTGCAGGACAACCCCGCGTTTCTCAA-+-------C 2001
          ||  |||  | ||||||   |   | | |   ||  ||||     | |          |
Db    621 CTCACCCAGACCTTCCCGGTGTCGCAGGAACGCACGCCCGTCGGCGGCGACCCGGCGCGC 562

Qy   2002 TTCCGCACCGAGGCCGGGCGCACGCTGTACGGCGAGGACGTCTACGTCGGGTACAGGTAC 2061
          |||| |  || || ||||     ||  |||   |||| | ||| |   || ||| | |
Db    561 TTCCCCGGCGTGGACGGGAAGGTGCACTACTCGGAGGGCATCTTCTCGGGCTACCGCTGG 502

Qy   2062 TACGAGTTTGCCGACAAGGACGTCAATTTCCCCTTTGGCCACGGCCTGTCCTACACCACT 2121
          |||||    |  |    ||||      |||||| || || ||||| || ||||||||||
Db    501 TACGACCGGGAGGGTGTGGACCCGCTTTTCCCGTTCGGGCACGGGCTTTCCTACACCACG 442

Qy   2122 TTTGCCTTTTCCAATCTCTCCGTGTCTCACAAGGACGGCAAGCTGAGCGTGTCCCTCTCC 2181
          || |     || | |   |||    | ||    || |   ||   |||| |  |  ||
Db    441 TTCGAGCGCACCGACCCGGTCGTCGAGCGCACCCGCGACGGCCTCGACGTGACGGTGACC 382

Qy   2182 GTGAAGAACACCGGCTCCGTGCCCGGCGCACAGGTGGCCCAGCTCTACGTCAAGCCCCTC 2241
          |||   ||||||||      || ||||    | || |  ||| ||||| ||  |||  |
Db    381 GTGCGCAACACCGGTCAGCGGCGCGGCAGCGACGTCGTGCAGGTCTACCTCGGCCCCAGC 322

Qy   2242 CAAGCGGCCAAGATTAACCGCCCCGTCAAGGAGCTCAAGGGCTTCGCAAAGGTCGAACTG 2301
          |    ||   | |  |||   |      | |||||   || | |   ||||||||||
Db    321 CCGCAGGTTCCGCTGGACCAGGCACCGCGGCAGCTCGCCGGGTACCAGAAGGTCGAACTC 262

Qy   2302 CAGCCCGGCGAGACGAAGGCGGTGACAATCGAGGAGCAGGAGAAGTACGTC 2352
            || |||||||| |||  ||     |  | |    ||||  |  | ||
Db    261 GCACCGGGCGAGACCAAGCGCGTCCGGGTGCACGTCGCGGAGCGGGCCCTC 211
```

Search completed: May 4, 2003, 03:05:15
Job time : 4797.77 secs